

Universidade de Lisboa
Instituto de Geografia e Ordenamento do Território



**Modelo geográfico de suporte à localização de estabelecimentos
comerciais de proximidade**

Ana Rita Ferreira da Silva de Almeida

Dissertação orientada pelo Prof. Doutor Fernando Jorge Pedro da Silva Pinto
da Rocha e coorientada pelo Dr. Celso Miguel Moreira Marques

Mestrado em Sistemas de Informação Geográfica e Modelação Territorial
Aplicados ao Ordenamento

2019



Modelo geográfico de suporte à localização de estabelecimentos comerciais de proximidade

Ana Rita Ferreira da Silva de Almeida

Dissertação orientada pelo Prof. Doutor Fernando Jorge Pedro da Silva Pinto da Rocha e coorientada pelo Dr. Celso Miguel Moreira Marques

Júri:

Presidente: Professor Doutor Herculano Alberto Pinto Cachinho do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa;

Vogais:

- Professor Doutor Nuno Manuel Sessarego Marques da Costa do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa;
- Professor Doutor José António Pereira Tenedório da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa;
- Professor Doutor Fernando Jorge Pedro da Silva Pinto da Rocha do Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa.

Agradecimentos

Em primeiro lugar tenho de agradecer ao Professor Doutor Jorge Rocha pela sua disponibilidade e amabilidade no decurso dos meses da realização da dissertação, e também um pouco depois disso. O professor foi o melhor orientador que poderia ter pedido, desde a sua ajuda na definição da dissertação até às suas tentativas (muitas vezes falhadas) de me tentar acalmar. Por isto, e muito mais, muito obrigada.

Em seguida, tenho de agradecer à minha família. À minha irmã por ser muito persistente para escrever a dissertação todos os dias, ao ponto de me fazer sentir mal por não estar a fazê-la, e à minha mãe por ser a calma durante a tempestade.

Por fim, aos meus amigos por me aturarem, por estarem lá quando mais preciso, e quando não preciso também.

Resumo

O setor de venda a retalho em Portugal, principalmente o alimentar, é um mercado muito competitivo, com elevados volumes de negócio, o que o torna num mercado muito atrativo. Ao longo do tempo, este setor foi dominado por diferentes fases de diversificação. A primeira fase marcada pelos estabelecimentos de proximidade explorados por pequenos retalhistas (década de sessenta); a segunda, a partir de meados da década de oitenta, com a entrada de novos operadores no mercado e a exploração de novos formatos (supermercados e hipermercados) que apostavam na acessibilidade; e por fim, devido ao crescimento contínuo e veloz da competitividade entre as organizações, e consequente saturação do mercado, as grandes distribuidoras voltam-se para os estabelecimentos de proximidade, recorrendo aos contratos de *franchising* para partilhar os riscos de exploração destas lojas.

A introdução da informação geográfica e aspetos espaciais nas empresas, ofereceram elementos indispensáveis à definição de estratégias. Com a utilização de dados provenientes de diversas fontes, as empresas conseguiram obter *insights* que não conseguiriam obter sem a componente espacial e puderam utilizá-los para otimizarem os seus negócios.

Esta dissertação tem como objetivo o desenvolvimento de uma proposta de localização para novos estabelecimentos comerciais de proximidade. Para isso pretende-se observar a distribuição dos estabelecimentos comerciais em Portugal Continental, com especial incidência sob uma das maiores cadeias de distribuição alimentar em Portugal; com a utilização de técnicas de cartografia dasimétrica pretende-se obter uma maior precisão na distribuição da população, de modo a definir mais concretamente a procura pelos estabelecimentos comerciais; assim como comparar as diferenças que os modelos estatísticos preditivos provocam na modelação do território e como estes vão influenciar na escolha das novas localizações para estabelecimentos comerciais de proximidade; e, por fim, com os modelos localização-alocação estudar-se-á a viabilidade das possíveis localizações para os novos estabelecimentos.

PALAVRAS-CHAVE: Localização; comércio de proximidade; Modelos localização-alocação; modelos preditivos; cartografia dasimétrica

Abstract

The retail sector, especially the food sector, is a very competitive market with high volumes of business, which makes it a very attractive market. Over time, this sector has been dominated by different phases of diversification. The first phase marked by proximity establishments operated by small retailers (sixties); the second, from the mid-eighties, with the entry of new operators into the market and the exploration of new formats (supermarkets and hypermarkets) that bet on accessibility; finally, due to the continuous and fast growth of competitiveness among organizations, and consequent saturation of the market, large distributors turn to local establishments, using franchising contracts to share the operating risks of these stores.

The introduction of geographic information and spatial aspects in companies provided essential elements for the definition of strategies. Using data from a variety of sources, companies were able to gain insights that they could not get without the spatial component and were able to use them to optimize their business.

This dissertation aims to develop a location proposal for new outlets. For this purpose, we intend to observe the distribution of commercial establishments in mainland Portugal, with special emphasis on one of the largest food distribution chains in Portugal; with the use of dasimetric mapping techniques, the aim is to achieve greater precision in the distribution of the population in order to define more precisely the demand for commercial establishments; as well as to compare the differences that the predictive statistical models provoke in the modelling of the territory and how these will influence in the choice of the new locations for commercial establishments of proximity; and, finally, with the location-allocation models, the viability of possible locations for the new establishments will be studied.

KEYWORDS: Location; Proximity retailing; location-allocation models; predictive models; dasymetric cartography

Índice geral

Agradecimentos	i
Resumo	iii
Abstract	v
Índice de figuras	ix
Índice de quadros	xi
Siglas/Acrónimos	xiii
1. Introdução	1
1.1. Objetivos	2
1.2. Metodologias	2
1.3. Estrutura da dissertação	3
2. Comércio de retalho	5
2.1. Marketing	9
2.2. Teorias de localização do comércio	11
2.3. Comércio alimentar de retalho em Portugal	15
2.4. Estabelecimentos de Proximidade	18
3. Geografia, SIG e <i>Geomarketing</i>	25
3.1. <i>Geomarketing</i>	26
3.2. Sistemas de Informação Geográfica (SIG)	28
3.3. SIG, <i>Geomarketing</i> e o comércio a retalho	30
3.4. Geodemografia	32
3.5. Decisão de localização do retalho	36
4. Informação de suporte à análise	39
4.1. Caracterização da Oferta	42
4.1.1. Difusão espacial	46
4.1.2. Caracterização das lojas Meu Super	49
4.2. Caracterização da Procura	55
4.2.1. Cartografia Dasimétrica	59
4.2.2. Geodemografia	62
4.3. Infraestruturas	73
4.3.1. Estabelecimentos	73
4.3.1.1. Densidade de lojas Meu Super	75
4.3.1.2. Densidade de lojas da concorrência	77
4.3.2. Densidade de estradas	78

4.3.3. Distância à concorrência	80
4.4. Áreas de Influência.....	80
4.4.1. Construção da rede	81
5. <i>Location Analytics</i>	85
5.1. <i>Data Mining</i>	85
5.2. Pré-processamento (Normalização das variáveis)	91
5.3. Modelos estatísticos	95
5.3.1. <i>Environmental Distance</i>	97
5.3.1.1. Distância euclidiana.....	101
5.3.2. <i>Support Vector Machines</i>	103
5.3.3. <i>Genetic Algorithm for rule-set prediction (GARP)</i>	106
5.3.4. Aptidão.....	109
5.3.1. Validação	111
5.4. Modelos de Localização-Alocação	112
5.4.1. Minimização da Impedância (p-mediana)	117
5.4.2. Maximizar a Cobertura	119
5.4.3. Minimizar as Instalações.....	120
5.4.4. Maximizar o Atendimento	122
5.4.5. Maximizar a Quota de Mercado	124
5.4.6. Capturar a Quota de Mercado	126
6. Considerações finais.....	129
7. Bibliografia	135
8. Anexos.....	147

Índice de figuras

Figura 2.1: Formas e formatos de retalho	7
Figura 2.2: Componentes do Marketing Mix	10
Figura 2.3: Modelos clássicos de Localização	12
Figura 2.4: Evolução do crescimento dos sectores de comércio	21
Figura 2.5: Funções da insígnia e do franqueado	22
Figura 3.1: Ordem cronológica do surgimento do <i>Geomarketing</i> , por ramos de conhecimento.....	26
Figura 3.2: Multidisciplinaridade do <i>Geomarketing</i>	27
Figura 3.3: Elementos do <i>Geomarketing</i>	31
Figura 3.4: Algumas aplicações comerciais da geodemografia	34
Figura 3.5: Fatores que afetam o sucesso de um estabelecimento de comércio de retalho alimentar.....	38
Figura 4.1: Esquema do processo de georreferenciação e métodos de Geocodificação. 40	
Figura 4.2: Esquema processual	41
Figura 4.3: Concorrência às lojas Meu Super, com área até 540 m ² , no ano de 2015....	42
Figura 4.4: Distribuição das lojas da Insígnia Meu Super, no ano de 2015.....	44
Figura 4.5: Distribuição Espacial da Concorrência. a) Pingo Doce; b) Lidl; c) Minipreço; d) Amanhecer	45
Figura 4.6: Escalas de distribuição de amostras	47
Figura 4.7: Distribuição espacial da loja Meu Super e da concorrência.....	48
Figura 4.8: Volume de vendas e transações das lojas Meu Super, de abril a maio de 2015.....	49
Figura 4.9: Lojas MS com menos transações comerciais trimestrais	50
Figura 4.10: Lojas MS com mais transações comerciais trimestrais	51
Figura 4.11: Lojas MS com menos vendas brutas trimestrais	52
Figura 4.12: Lojas MS com mais vendas brutas trimestrais	52
Figura 4.13: Média de Gastos, com cartão, por município	54
Figura 4.14: Informações retiradas dos inquéritos realizados pela Sonae em 2015	56
Figura 4.15: Resultados dos inquéritos	56
Figura 4.16: Modelo de cartografia dasimétrica.....	61
Figura 4.17: Exemplo da distribuição dasimétrica da população	62
Figura 4.18: Distribuição do número de indivíduos residentes, em 2011	64
Figura 4.19: Distribuição do número de mulheres (a) e homens (b) residentes, em 2011	65
Figura 4.20: Distribuição do número de indivíduos residentes entre os 20 e os 64 anos (a) e com mais de 64 anos (b), em 2011	66
Figura 4.21: Distribuição do número de famílias clássicas compostas por 1 ou 2 elementos, em 2011	67
Figura 4.22: Distribuição do número de residentes empregados (a) e pensionistas ou reformados (b), em 2011	68
Figura 4.23: Distribuição do nº de Indivíduos Residentes empregados no sector primário (a); Nº Indivíduos Residentes empregados no sector secundário (b); Nº Indivíduos Residentes empregados no sector terciário (c), em 2011	70
Figura 4.24: Distribuição do nº de Indivíduos Residentes com o 1º Ciclo (a); Nº Indivíduos Residentes com o 2º Ciclo (b); Nº Indivíduos Residentes com o 3º Ciclo (c);	

Nº Indivíduos Residentes com o Ensino Secundário (d) e Nº Indivíduos Residentes com um Curso Superior, , em 2011. Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI).....	72
Figura 4.25: Densidade de lojas MS, por vendas brutas	76
Figura 4.26: Densidade de lojas de insígnias concorrentes, por área de loja	77
Figura 4.27: Densidade de estradas	79
Figura 4.28: Exemplo das áreas de influência das lojas Meu Super.....	84
Figura 5.1: Etapas do processo de Extração de Conhecimento em Bases de Dados	86
Figura 5.2: Correlação entre as variáveis	90
Figura 5.3: Variáveis escolhidas para os modelos (resultado do VIF)	91
Figura 5.4: Lógica Booleana (a) vs. Lógica <i>Fuzzy</i> (b).....	93
Figura 5.5: Exemplo do ficheiro de ocorrências (nota: as casas decimais estão representadas por pontos devido às especificidades do <i>software</i>).....	97
Figura 5.6: Algoritmo Environmental Distance	99
Figura 5.7: Probabilidade do modelo	100
Figura 5.8: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a aplicação algoritmo <i>Environmental Distance</i> (através da distância euclidiana)	102
Figura 5.9: Hiperplanos das SVM	104
Figura 5.10: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a aplicação do algoritmo <i>Support Vector Machines</i>	105
Figura 5.11: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a aplicação do algoritmo GARP.....	108
Figura 5.12: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a média dos algoritmos aplicados	110
Figura 5.13: Lojas encerradas com aptidão elevada	112
Figura 5.14: Resultado do modelo de minimização da impedância	118
Figura 5.15: Resultado do modelo de maximização da cobertura.....	120
Figura 5.16: Resultado do modelo de minimização das instalações.....	122
Figura 5.17: Resultado do modelo de maximização do atendimento	123
Figura 5.18: Resultado do modelo de maximização da quota de mercado	125
Figura 5.19: Resultado do modelo de Target Market Share, referente ao ano de 2017	127

Índice de quadros

Quadro 2.1: Empresas por Dimensão de Pessoas ao Serviço.....	18
Quadro 3.1: Principais definições de SIG	29
Quadro 4.1: Dados utilizados	39
Quadro 4.2: Número de Lojas de cada insígnia em Portugal Continental	43
Quadro 4.3:Características do padrão de distribuição da loja Meu Super e da concorrência	48
Quadro 4.4: Dados provenientes dos censos 2011	58
Quadro 5.1: Funções <i>Fuzzy</i> (ArcGIS help)	94
Quadro 5.2: Aptidão das lojas encerradas - validação	111
Quadro 5.3: Funções- objetivo dos modelos localização-alocação (ArcGIS help)	116
Quadro 8.1: Dados financeiros das lojas Meu Super	147
Quadro 8.2: Matriz de Correlação	147

Siglas/Acrónimos

AMA – *American Marketing Association*

CAOP – Carta Administrativa Oficial de Portugal

DDP – Distribuição Dasimétrica da População

DP – Distribuição Potencial (modelo DP)

ECBD - Extração de Conhecimento em Bases de Dados

ED – *Environmental Distance*

GARP - *Genetic Algorithm for rule-set prediction*

IVMP – Índice do Vizinho Mais Próximo

MS – Meu Super

oM – openModeller

SIG – Sistemas de Informação Geográfica

SVM – *Support Vector Machines*

u.m – Unidades Monetárias

1. Introdução

A geografia desempenha um papel fundamental no sucesso de um negócio (Alcaide, Calero, & Hernández, 2012; García-Palomares, Gutiérrez, & Latorre, 2012, citados por Roig-Tierno et. all, 2013). No retalho, é importante ter uma estratégia de localização adequada, portanto é crucial realizar uma análise sólida dos possíveis locais para novas aberturas de lojas. Neste setor, a abertura de uma nova loja é uma decisão crítica, pois a escolha do local implica sérios riscos financeiros e na imagem corporativa para a empresa em questão.

De acordo com Cliquet (2006), os estudos em *Geomarketing* estão assentes em três grandes blocos: técnicas de codificação, delimitação geográfica e funções estatísticas, que em conjunto permitem manipular os dados de forma a compreender e planear o território. A introdução da informação geográfica e aspetos espaciais nas empresas, ofereceram elementos indispensáveis à definição de estratégias. Com a junção da Geografia ao Marketing (*Geomarketing*), as empresas têm conseguido potenciar os seus recursos, uma vez que apostando no estudo da localização dos seus serviços as empresas poderão assegurar o seu sucesso (Simmons & Jones, 1990, citados por Afonso, 2018), visto que uma boa localização pode permitir uma redução de custos e uma melhoria na acessibilidade, assim como obter uma vantagem sobre a sua concorrência.

O sector do comércio de retalho é conhecido por ter elevados valores de densidade por todo o território (Dias, 2013, p.6), uma vez que é um sector fortemente implantado no território. Com o aparecimento de novos formatos de comércio de retalho e, conseqüentemente, o aumento da concorrência entre eles, torna-se imperativo que as empresas comecem a prever a penetração de mercado por parte dos concorrentes, e de se anteciparem através da fixação de várias lojas em diferentes localizações (Cliquet, 2006), sendo necessário encontrar as localizações ótimas e determinar a alocação da procura (Buzai, 2013).

Planear a localização de equipamentos do sector privado, devido aos custos financeiros inerentes, e tendo como objetivo da maximização da sua eficácia, implica a escolha de critérios que fundamentem e apoiem inequivocamente a tomada de decisão. Para se escolher o melhor local é necessária uma análise detalhada das características da comunidade a quem vai servir e do território onde se vai inserir. Deste modo, os SIG conseguiram processar grandes volumes de informação, e gerar modelos de otimização

de localização, em curto espaço de tempo, facilitando o processo de decisão e na concepção de soluções, por parte do decisor.

1.1. Objetivos

Os objetivos a alcançar prendem-se com o desenvolvimento de uma proposta de localização para novos estabelecimentos comerciais de proximidade, sustentada na cartografia dasimétrica para o cálculo preciso da distribuição da população, no cálculo de áreas com potencial de aptidão e na aplicação de modelos de redes.

Para alcançar os objetivos desta dissertação pretende-se:

1. Discutir os conceitos de comércio a retalho e comércio de proximidade, mediante a leitura ativa de artigos científicos, de modo a enquadrar o tema da dissertação;
2. Definição das áreas de procura dos estabelecimentos Meu Super através do uso da cartografia dasimétrica;
3. Analisar a distribuição dos estabelecimentos comerciais, através de técnicas baseadas em *Geomarketing*, para encontrar padrões na distribuição da oferta comercial;
4. Avaliar as potencialidades e limitações dos diferentes modelos espaciais de modo a encontrar o modelo síntese de aptidão;
5. Identificar as localizações ótimas para a abertura de novos estabelecimentos comerciais de proximidade através de modelos localização-alocação.

1.2. Metodologias

É no contexto de aumento da concorrência entre as grandes distribuidoras e a aposta nos estabelecimentos de proximidade que esta dissertação propõe encontrar as localizações ótimas para a abertura de novos estabelecimentos comerciais de proximidade Meu Super. Para isso, foi necessário localizar as lojas existentes, quer da insígnia em estudo como as suas concorrentes, mediante a geocodificação de endereços; conhecer o perfil de cliente e definir as áreas onde estes se encontram, através da cartografia dasimétrica; criar a *Network Dataset* e as restantes variáveis através da densidade de

Kernel e da distância euclidiana. Com as variáveis criadas procedeu-se, com o auxílio do *software* OpenModeller, à realização de três algoritmos para o cálculo de aptidão (*Support Vector Machines, Environmental Distance, Genetic Algorithm for rule-set prediction*). Estes modelos foram escolhidos de modo a demonstrar os três níveis de dificuldade dos algoritmos presentes neste *software*. Por fim, com as melhores localizações, provenientes da média dos três algoritmos, e com o apoio da *Network Dataset* criaram-se os modelos localização-alocação, para assim se obter as melhores localizações para a abertura dos novos estabelecimentos.

1.3. Estrutura da dissertação

Esta dissertação será organizada em quatro capítulos, excluindo os capítulos que dizem respeito à introdução e conclusões. Seguidamente, apresentar-se-á uma breve apresentação dos mesmos.

No segundo capítulo, é feita a primeira parte da revisão da literatura, abordando-se o comércio a retalho, as suas formas e formatos. Começa-se por introduzir o conceito de comércio a retalho e como a localização é importante para determinar o sucesso das empresas, indicando as principais teorias que introduziram a localização nas teorias económicas. Em seguida, aborda-se o comércio a retalho em Portugal, com especial incidência no comércio de retalho alimentar.

No terceiro capítulo é apresentada a segunda parte da revisão bibliográfica, abordando a evolução do *Geomarketing*, começando com o surgimento dos sistemas de informação geográfica e, posteriormente, do *Geomarketing* e da geodemografia. Como estas ciências são importantes no comércio a retalho, e como se processa a decisão da localização de novos estabelecimentos comerciais de proximidade.

No quarto capítulo é apresentado o material utilizado e descrevem-se, de forma pormenorizada, o tratamento dos dados e da metodologia de processamento.

Por fim, no quinto capítulo são apresentadas novas metodologias para a realização dos modelos de aptidão, e é feita a apresentação dos resultados, a análise e discussão dos mesmos. Faz-se a análise comparada dos resultados obtidos nos diferentes métodos e, em seguida, realizam-se os modelos localização-alocação para mostrar as melhores localizações para a abertura de novas lojas Meu Super (MS).

2. Comércio de retalho

A venda a retalho é um dos maiores setores do mundo (Zentes et al., 2007, p.1) e um dos sectores mais dinâmicos e disputados na economia (Cachinho, 1999, p.98). Este, representa a fase final da distribuição das mercadorias, competindo aos retalhistas a venda das mercadorias em quantidades mais reduzidas ao público.

A palavra “retalho”, provem da palavra francesa “*tailer*” que significa “cortar, aparar, dividir”, utilizado na costura. Apenas em 1433, foi aplicado, pela primeira vez, com o significado de “venda em pequenas quantidades” (RAI, sem data; Santos, 2014). Segundo Berman & Evans (1995), retalho refere-se à venda fracionada de artigos em quantidades adequadas às necessidades do consumidor final, tendo o mesmo significado em francês, alemão e holandês.

A atividade comercial tem vindo a acompanhar a evolução da população, adaptando-se às necessidade e aos hábitos de consumo desta ao longo do tempo (Cachinho, 1999). O comércio está na vida da população desde o surgimento das primeiras cidades no Egipto e na Mesopotâmia no séc. IV a.C., onde existia um espaço público para a troca de produtos (Benévolo, 1993). Com o passar dos anos, a atividade comercial alterou-se, sendo que as principais diferenças passam a ser entre os bens que se trocavam e o local onde as transações se fazem.

No século XX assistiu-se ao surgimento e desenvolvimento de grandes indústrias de bens de consumo não-duráveis, *e.g.*, a Unilever, Colgate, Johnson & Johnson e a Coca-Cola, que consolidaram a sua presença em diversos países, através do comércio de retalho, e na mente dos consumidores com as suas estratégias de marketing (Motta & Silva, 2006).

No fim do século passado, o comércio de retalho foi uma das áreas que impulsionaram a economia global, assistindo-se a elevadas taxas de crescimento nos países desenvolvidos e crescimento exponencial equilibrado nas economias emergentes (Zulqarnain et al., 2015). Este crescimento rápido, apesar de ser um fator importante, foi seguido pelo aumento da concorrência e pelo surgimento de novos formatos de retalho e tipologia de lojas (Leszczyc & Timmermans, 1997). O comércio a retalho possui uma ampla variedade de modelos e atende às variadas necessidades da população, sendo composto por diversos formatos de lojas que diferem segundo o seu tamanho e a variedade de bens comercializados (Parente, 2000, citado por Motta & Silva, 2006).

Hoje em dia, não existe uma definição global de retalho uma vez que cada autor modifica a definição para se adaptar ao seu trabalho. Por exemplo, Kotler (1997) define a atividade de retalho como a venda ou comercialização de produtos e serviços, em pequenas quantidades, para, deste modo, conseguir chegar ao consumidor final; Rousseau, em 2008, define retalho como a venda de bens e serviços a consumidores finais, onde se adaptam os produtos em quantidades unitariamente de acordo com o poder aquisitivo e à satisfação das necessidades dos consumidores; e ainda o Decreto-Lei nº 339/85 de 21 de Agosto, da lei portuguesa, define comércio a retalho como “... toda pessoa física ou colectiva que a título habitual e profissional, compra mercadorias em seu próprio nome e por sua própria conta e as revende directamente ao consumidor final” (Ministério do Comércio e Turismo, 1985). Apesar das várias definições, é seguro afirmar que o retalho consiste na compra de produtos, por parte de uma determinada loja, a um grossista, com o objectivo de vender ao consumidor final, um determinado número de bens a preços competitivos e acessíveis.

Já em 1971, Markin, citado por Santos (2014), afirmava que o comércio de retalho era tão omnipresente no quotidiano, que era difícil não ter contacto com alguma das suas manifestações durante o dia. Atualmente, é praticamente impossível. Este sector tem diversas formas de organização de negócio, cada uma com a sua peculiaridade e com diferenças ao nível dos consumidores atingidos, forma de atuar no mercado e rentabilidade.

Segundo Rousseau (2008, p.61), o retalho apresenta três formas (Figura 2.1):

- Retalho independente, que se refere a pequenos estabelecimentos de natureza artesanal e/ou familiar de dimensões reduzidas, constituído por apenas um ponto de venda (*e.g.* retalho de proximidade (mercearias e drogarias), retalho especializado (padarias e talhos) e o retalho não sedentário (feiras de rua);
- Retalho integrado, que abrange as empresas comerciais que pertencem ao mesmo grupo económico e que fazem as funções de grossistas e retalhistas e, por vezes funções de produção, distribuição e venda (*e.g.* a Modis actua como plataforma logística de todas as insignias do Grupo Sonae – Continente, Worten, MeuSuper, etc.);
- Retalho Associado, que é formado por empresas com independência jurídica que se associam para realizar operações comuns de compras e serviços de modo a

obter economias de escala e tentar competir com o comércio integrado. Este pode ser distinguido em duas categorias, as cadeias voluntárias, em que o comerciante grossista tenta fidelizar os retalhistas oferecendo melhores condições em troca de quotas de compra (*e.g.* o Recheio e os estabelecimentos Amanhecer); e as cooperativas retalhistas, onde os retalhistas se organizam para assumirem as funções dos grossistas, fazendo com que os cooperantes consigam os benefícios económicos e logísticos que de outra forma não conseguiriam (*e.g.* a cooperativa Gct e as insignias Ponto Fresco, Coviran, Intermaché e SPAR).



Figura 2.1: Formas e formatos de retalho

Fonte: Rousseau (2008, p.61-133)

De acordo com José António Rousseau, entrevistado por Victor Jorge, em 2009, existe cada vez mais concorrência no mercado do comércio a retalho, havendo a preocupação, por parte dos grandes retalhistas, de adquirirem a maior quota de mercado possível e, ao mesmo tempo, oferecer uma vasta gama de produtos a preços competitivos e inovadores, para que os consumidores optem pela sua loja em função das outras.

Com a modernização do sector do retalho a partir dos anos sessenta (anos oitenta, em Portugal) e com o aumento do poder de compra por parte da população, por exemplo, através do aumento dos rendimentos com o crescimento do emprego feminino (Pacione,

2005), os consumidores tiveram à sua disposição uma oferta diversificada de bens a preços competitivos. Devido ao aumento do rendimento disponível para o consumo, houve um aumento do número de estabelecimentos que o consumidor tem à sua disposição. Com o aumento da concorrência houve a necessidade de os retalhistas estarem em constante adaptação e de melhorarem os seus serviços. Dependendo do tipo de cliente, Lambin (1995) citado por Castro (2007, p.8), afirma que existem seis atributos a ter em consideração por parte de um ponto de venda ou unidade comercial:

- A proximidade: que irá definir a área de influência onde o retalhista irá estabelecer relações comerciais com os seus clientes;
- A variedade da oferta de produtos;
- O nível de preços praticados: compreende a amplitude de preços praticados, a presença regular de promoções, reduções de preços por quantidade, etc.;
- O nível de serviços: compreende serviços de conforto (parque, assistência, espaços de descanso, espaços diversão infantil...), serviços técnicos (método de venda, assistência pós-venda, entrega, montagem e reparação, ...), serviços financeiros (aceitação de cheques, cartões de crédito, financiamento, ...) e serviços paracomerciais (agências de viagens, stands especializados, listas de ofertas, ...);
- O tempo: facilidades de acesso, facilidade na identificação das secções dentro da loja, ausência de filas nas caixas e toda a organização interna da loja;
- O ambiente: a luminosidade, ambiente musical, ausência de ruídos perturbadores e a decoração interior.

De acordo com Berman e Evans (1995), o comportamento dos clientes varia com o local onde se fazem as compras e com o nível de envolvimento no ato de comprar. Estes critérios são utilizados pelos consumidores como critérios de comparação entre diferentes unidades comerciais concorrentes, fazendo com que haja a necessidade, da parte do retalhista, em ponderar as suas escolhas e definir um conceito que atraia os consumidores à sua loja. Para isso poderá recorrer a estratégias de marketing de forma a diferenciar-se da concorrência.

2.1. Marketing

Schewe e Smith (1982) situaram a origem do marketing nos anos 50, a partir do momento em que os fabricantes se aperceberam que a sua produção devia considerar as necessidades dos consumidores, de modo a dar resposta a essas necessidades. A partir dessa década, começou a haver, por parte das empresas, uma maior preocupação relativamente à melhoria da distribuição de bens e serviços, o que introduziu a necessidade de aplicar uma visão geográfica ao marketing.

Atualmente, não existe uma definição geral de Marketing, mas ao longo do tempo, foram vários os autores que teceram as suas considerações relativamente à definição de marketing. Levitt, em 1960, afirma que o marketing visa a satisfação das necessidades do cliente através do produto e do conjunto de fatores associados à sua criação, entrega e consumo. Em 2003, Kotler, introduz o Marketing como a arte e a ciência de escolher os mercados alvo e de conquistar, reter e cultivar clientes, através da criação, comunicação e fornecimento de valor superior para os mesmos. E mais recentemente, a AMA (American Marketing Association, 2013), que a cada cinco anos atualiza a definição de Marketing, definiu Marketing como a atividade, o conjunto de instituições e processos, para criar, comunicar, entregar e trocar ofertas que tenham valor para consumidores, clientes, parceiros e sociedade em geral.

Os retalhistas são o elo entre os fabricantes e os consumidores, apresentando-se como intermediários no canal de distribuição. No entanto, embora, por vezes, considerado um elo passivo no canal, os retalhistas estão a aproveitar a sua posição para se tornarem o ator dominante no canal de distribuição (Zentes et al., 2007). Deste modo, nos últimos anos tem-se assistido a um crescimento da importância do Marketing nas organizações retalhistas, muito devido ao aumento da concorrência, com o aparecimento dos *hard discounts*; à globalização dos formatos retalhistas; das parcerias entre organizações de modo a satisfazer os clientes a um menor custo; das melhorias competitivas; e com o crescimento das marcas próprias dos retalhistas em detrimento de marcas dos produtores (Collins e Burt, 2006 citados por Castro, 2007), fazendo com que este permita que hoje o comércio não se limite a satisfazer as necessidades e desejos do consumidor, mas participe diretamente na formação dessas necessidades (Cachinho, 1999).

No âmbito do marketing estratégico de retalho, os retalhistas têm mais opções disponíveis no seu *marketing mix* do que os fabricantes, uma vez que estes estão em

contacto direto com os consumidores finais, que visitam e interagem diretamente com eles (Zentes et al., 2007). Com as variadas exigências dos consumidores e o aumento da concorrência, as lojas necessitam de obter vantagem sobre os seus concorrentes diretos. Nesta fase, os retalhistas necessitam de investir no marketing e ter especial atenção a todos os componentes do *Marketing Mix* (Figura 2.2). O *Marketing Mix* refere-se a um conjunto de ações ou táticas que uma empresa usa para promover a sua marca ou produto no mercado. Tipicamente, este é composto por 4P's: *Price*, *Product*, *Promotion* e *Place* – ou seja – Preço, Produto, Promoção e Localização. No entanto, hoje em dia, este *mix* inclui cada vez mais elementos vitais (mais P's), *e.g.*, as pessoas (*people*), o posicionamento (*positioning*), o acondicionamento (*Packaging*) e até mesmo a política (*politics*) (The Economic Times, sem data).

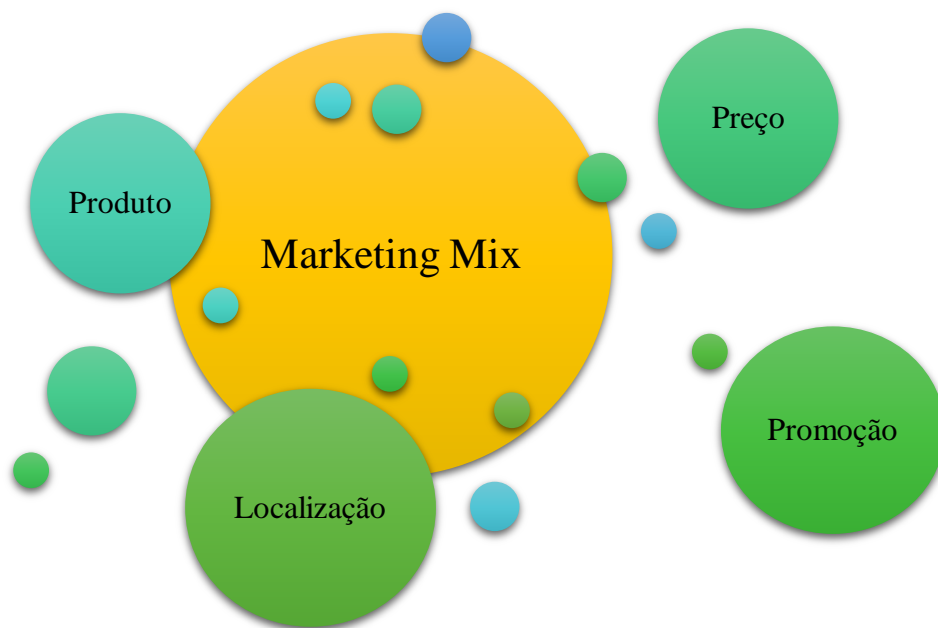


Figura 2.2: Componentes do Marketing Mix

Todos os elementos do *marketing mix* se influenciam mutuamente, sendo que estes são essenciais a um bom plano de negócio de uma empresa. Estes componentes bem utilizados e pensados podem levar ao sucesso de um negócio, ou se tratados sem atenção podem prejudicar um negócio que poderá não recuperar. Quando se está a estabelecer um negócio é necessário ter-se em atenção (The Economic Times, sem data):

- Os produtos a vender, que são a resposta que a empresa oferece para suprimir as necessidades do público-alvo. Segundo Kotler e Armstrong (2003) os aspetos mais relevantes destes são a tangibilidade, a durabilidade e o preço;
- O preço dos produtos, que irá depender dos custos de produção, do segmento de mercado, do fator procura-oferta e de uma série de outros fatores diretos e indiretos. O preço também pode ser usado para diferenciar e melhorar a imagem de um produto;
- A promoção, que se refere a todas as atividades realizadas para tornar o produto ou serviço conhecido para os consumidores (*e.g.*: publicidade);
- A localização, que se refere ao ponto de venda. Este é um dos fatores essenciais para o sucesso de um negócio, uma vez que atrair a atenção dos consumidores é o objetivo principal de todas as empresas, e apostar numa boa localização é chave.

Desde os primeiros modelos de localização do comércio, vários autores afirmaram que a localização é essencial ao sucesso de um retalhista, entre eles Jones e Simmons (1990) que no livro “The retail environment”, afirmam que existem três fatores essenciais para o sucesso de um retalhista “*Location, location, location*”, ou seja, “Localização, localização, localização”. Para este e mais autores, como Kotler e Armstrong (2003, p.346) e Zentes et al. (2007, p.4), a localização do ponto de venda é considerado um fator determinante no sucesso de uma loja, uma vez que influencia diretamente na capacidade da empresa de atrair clientes e, por outro lado, os custos das instalações também causam impactos significativos no resultado financeiro das empresas.

2.2. Teorias de localização do comércio

Para uma empresa do sector do retalho, é da maior importância retirar o melhor proveito da localização, uma vez que uma boa localização pode contribuir para o sucesso de uma empresa. Este facto coloca a Geografia como uma das ciências mais inseridas no domínio empresarial (Cliquet, 2006). A compreensão da localização das atividades económicas é um tópico de pesquisa que tem suscitado muito interesse por parte dos investigadores ao longo de alguns séculos (Figura 2.3).

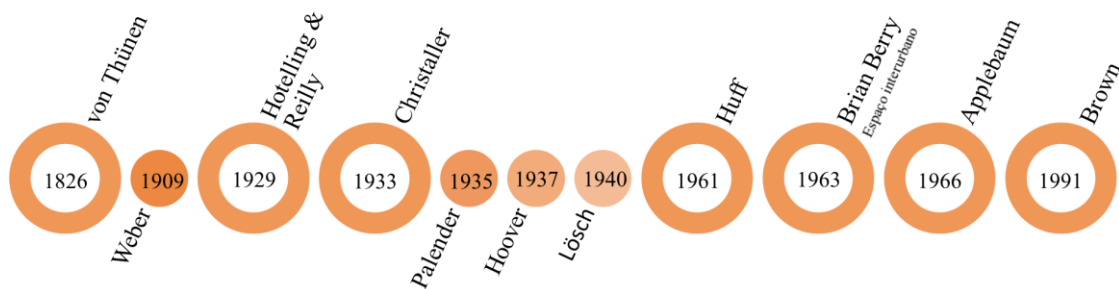


Figura 2.3: Modelos clássicos de Localização

As origens mais remotas da análise económica espacial datam o final do século XIX e princípios do século XX, quando um grupo de investigadores começou a desenvolver essa linha de pensamento (Yrigoyen, 2003). O responsável pela introdução da dimensão espacial nas teorias económicas foi Johann Heirich von Thünen no seu livro *Der Isolierte Staat* (O Estado Isolado), de 1826, onde deu a conhecer o seu modelo de localização agrícola, em que propunha a interpretação da análise económica sobre o espaço geográfico com base na renda da terra (valor), na distância dos agricultores ao mercado (custos de transporte) e nos preços aplicados pelos agricultores (preços dos produtos) (Bradford & Kent, 1987), estudando a localização de diferentes culturas em torno dos centros urbanos. Alfred Weber, seguiu os ensinamentos de von Thünen, mas aplicados à localização industrial, apresentando uma teoria mais realista, pois a distribuição espacial dos fatores de localização apresenta-se de forma discreta sobre o território, possibilitando para além de explicar e descrever, estipular a localização das atividades. No entanto, nenhuma destas teorias tiveram em consideração a competição entre organizações em respeito ao seu comportamento locativo.

A primeira teoria a considerar a concorrência foi a de Hotelling (1929) que afirmou que existe uma “tendência para competidores imitarem-se uns aos outros em termos de qualidade de bens, localização e em outros fatores essenciais”. A teoria com o seu nome, também conhecida como a teoria dos princípios mínimos de diferenciação, explica por que razão os restaurantes e as lojas de retalho se costumam localizar perto uns dos outros, utilizando o exemplo de dois vendedores de gelados perto um do outro na praia. Hotelling foi um dos primeiros economistas a assumir que existe interdependências entre empresas e que a localização de uma afeta as restantes, especialmente, no que se refere às áreas de mercado, uma vez que serão competidores.

Em 1929, William Reilly apresentou os Modelos Gravitacionais, demonstrando a utilidade dos modelos de Issac Newton no Marketing Geográfico, com a sua “Lei da Gravitação do Retalho”, onde afirmava que a interação potencial entre consumidores e vários pontos retalhistas dentro de uma área urbana varia diretamente com o poder de atração (ou tamanho) de cada ponto, e inversamente em relação à distância que separa o consumidor do ponto de origem (Machado et al., n.d.), ou seja, propôs que os consumidores consideravam, não só a distância, mas também a atratividade das diferentes alternativas de retalho.

Considerando o tema desta dissertação, a teoria mais relevante é a “Teoria dos lugares centrais” de Christaller (em 1933), visto que esta foi considerada de grande importância para o desenvolvimento do *Geomarketing*, procurando organizar a complexidade relativa às decisões espaciais do comércio. Christaller procurou compreender as leis que determinavam o tamanho e a distribuição das cidades, propondo que o espaço se organizava através de um padrão de centros hierarquizados, diferenciados pela sua força de atração em relação às áreas vizinhas. De acordo com Cavalcante (2005), citado por Furlan (2011), Christaller admite que a produção de bens e serviços nas cidades resultaria de uma escala de produção que alcança um ponto ótimo, representado pela procura dividida num espaço homogêneo. Ao aplicar o seu método, o autor conclui que haveria uma tendência para formação de arranjos específicos (em forma hexagonal) para a distribuição das cidades numa determinada região.

Retomando a linha de pensamento desenvolvida por Weber, o sueco Palander (1935) tentou elaborar uma teoria geral de equilíbrio espacial para duas empresas. Este concentrou-se em estudar os efeitos dos preços sobre as extensões de mercado em que as empresas podem vender os seus produtos quando a localização, as condições de concorrência, os custos dos fatores e as tarifas de transporte são fornecidos. Hoover (1937) alargou o exemplo de Palander para ser utilizado para mais que duas empresas, providenciando um exemplo gráfico útil da derivação das áreas de mercado para três empresas em competição, utilizando um método de análise que obviamente podia ser extensível a mais casos. Considera-se que cada uma das empresas está otimamente localizada no sentido Weberiano do termo, em termos de entradas (e é neste sentido que os seus custos de produção podem diferir, *e.g.* o custo de mão-de-obra pode ser diferente ou a matéria-prima pode estar a distâncias diferentes) (Martins, 2012). Hoover desenvolveu, assim, um modelo que relaciona a procura espacial e a receita marginal,

demonstrando que há uma tendência crescente de preços quando os custos da unidade de transporte crescem, introduzindo a análise da discriminação de preços espaciais (Duch, 2005).

Lösch (1940) teceu críticas tanto à teoria de von Thünen como a de Weber, assumindo que a escolha do local deve procurar o maior lucro possível e não o menor custo possível. A sua principal preocupação era desenvolver um modelo de equilíbrio geral do espaço, que servisse como orientação básica para o planejamento eficiente, tanto do ponto de vista privado, quanto público, focando a maximização de lucros como fundamento para a escolha da localização ótima. Lösch introduziu a procura na teoria da localização, admitindo que “a elasticidade da curva da procura varia segundo as regiões, conforme a preferência dos consumidores e, principalmente, de acordo com o seu poder aquisitivo” (Manzagol, 1985, p. 34, citado por Júnior, 2002, p.34).

O primeiro modelo que considerou a preferência dos consumidores foi desenvolvido por Huff (1964). Este usou as obras de Reilly e reformulou o modelo introduzindo a complexidade das necessidades dos clientes no estudo do *Geomarketing* (Bernardes, 2017). De acordo com Cliquet (2006), Huff determinou que a dimensão de uma loja deveria ser considerada como um indicador de atratividade, visto que um consumidor estará disposto a deslocar-se mais quanto mais produtos oferecer um estabelecimento, ou seja, quanto maior um estabelecimento comercial. Huff sugeriu, assim, que a atratividade de uma loja varia proporcionalmente à dimensão da loja e inversamente proporcional à distância.

Brian Berry (1963) veio revolucionar a forma como se entendia a organização dos espaços comerciais no espaço interurbano. Este repensa a teoria de Christaller e aplica-a ao espaço interurbano, acrescentando à estrutura dos centros hierárquicos (centros de conveniência/proximidade ao CBD), as faixas comerciais (*e.g.*: Estrada de Benfica) e áreas especializadas (*e.g.*: zonas de lazer noturno, como o Bairro Alto) (Salgueiro & Cachinho, 2011). Esta teoria revela-se de grande importância para o tema em estudo, visto que introduz os centros de proximidade nas áreas interurbanas, áreas onde se encontram as lojas de proximidade estudadas.

Applebaum (1966) criou o primeiro modelo sistemático de predição de retalho com base em dados empíricos, conhecido como método análogo. Este modelo consistia na utilização de dados de lojas existentes, como dados análogos, para a previsão de vendas

de uma instalação similar. Utilizava dados como origens geográficas, características demográficas e hábitos de consumo dos clientes de lojas análogas para determinar, através do mapeamento, a área de influência primária de um conjunto de lojas. Em seguida, cruzava estas informações com as vendas reais das lojas para avaliar as vendas de possíveis localizações futuras. Applebaum considerava que se a área de influência estimada de uma futura localização tinha um número de pessoas residentes e de concorrência similar a de lojas existentes, então poderiam ser feitas comparações para se calcular as vendas da nova localização (Aranha & Figoli, 2001).

Por último, Brown (1991) acreditava que a teoria de Berry carecia de universalidade e anacronismo. Assim, tentando afastar-se das hierarquias proferidas na teoria dos lugares centrais, Brown admite não haver hierarquias, uma vez que o comércio apresenta várias vertentes diferentes, fazendo com que seja mais complicado hierarquizá-las. Por isso define uma matriz composta de nove categorias de áreas de venda a retalho, em que se cruzam a forma das áreas de venda a retalho (agrupada, linear e isolada) com as funções dos espaços e lojas (generalizadas, especializadas e auxiliar). (Salgueiro & Cachinho, 2011). Assim, Brown acreditava que a atratividade de uma loja não é definida pelo local onde está inserida ou pela sua dimensão, mas pelas funções oferecidas.

Estas foram as teorias mais relevantes no que se refere à localização das atividades económicas e as que estiveram na origem do *Geomarketing*. Estas teorias revelaram-se corretas, mas generalistas, uma vez que cada cidade tem a sua realidade e em cada país a sua realidade é diferente e os retalhistas têm de se adaptar a essa realidade.

2.3. Comércio alimentar de retalho em Portugal

O comércio a retalho apresenta uma grande variedade de formas (estabelecimentos, comércio electrónico, feiras, ...), formatos (mercearia tradicional, mercearia com livre serviço/minimercado, supermercado, estabelecimentos especializados, hipermercados), localizações (urbana – zona comercial, zona residencial, zona mista – periferia, vila/aldeia), produtos (alimentares, não alimentares, medicamentos, ...), estruturas jurídicas (em nome individual, sociedades anónimas, sociedade quotas, cooperativa), entre outros (Salgueiro et al., 2002, p304-305; Fernandes et al., 2000, p.11). Este projecto assentará no comércio alimentar, sendo neste domínio em que se focará a análise.

Ao longo dos anos, um pouco por todo o mundo, assistiu-se a grandes mudanças no sector do retalho alimentar, entre elas a globalização e consolidação das grandes redes de supermercados (*e.g.*: Carrefour); o surgimento de novos formatos de retalho (*e.g.*: lojas de descontos); difusão de marcas próprias; constituição de centrais de negócios por parte das grandes distribuidoras (*e.g.*: compra dos supermercados Modelo por parte do grupo Jerónimo Martins); aumento da concorrência entre retalhistas de configurações distintas; e mais recentemente, a implementação do *e-commerce*/retalho *on-line* (Motta & Silva, 2006).

De acordo com Santos (2014, p.17), o comércio alimentar apresenta quatro tipologias:

- Tradicional, constituído por todos os exemplos de pequena distribuição em que a estrutura accionista é muito pequena, sendo maioritariamente constituída por um dono;
- Supermercado, que o Food Market Institute (2018) define como uma loja alimentar em livre serviço que possui um stock entre 15 000 e 60 000 produtos, e um volume de vendas de dois milhões de dolares anuais;
- *Discount*, caracterizada pelo auto-serviço que procura retirar mais custos de exploração da atividade, o escasso número de trabalhadores e um reduzido número de itens (600 a 1000). Estas lojas costumam ter pequenas áreas de venda (400 a 1000 m²) e o sortido é formado essencialmente por marcas do distribuidor, permitindo adoptar uma política de preços entre 20% e 40% inferior aos preços de mercado;
- Hipermercados, trata-se de uma fórmula comercial que combina o supermercado alimentar e a *discount store* não alimentar. A área de venda mínima destas superfícies varia de país para país, sendo que em Portugal o mínimo é de 2000 m² e costuma pertencer à grande distribuição.

No aparelho comercial português encontra-se duas vertentes distintas de comércio. O tradicional, consituído por pequenos estabelecimentos, generealistas ou especializados na oferta de um número reduzido de produtos afetos a ramos de atividade específicos; e uma vertente mais moderna, decorrente da revolução comercial, que se

caracteriza pela diversidade de formatos e pela grande dimensão de algumas unidades (*e.g.* hipermercados)

O primeiro hipermercado inaugurado em Portugal, mais precisamente, em Matosinhos, foi um Modelo Continente do grupo Sonae a 10 de Dezembro de 1985. Anteriormente, o comércio em Portugal era maioritariamente tradicional, com as lojas de bairro e alguns supermercados. Desde então, houve uma grande evolução no sector da distribuição em Portugal, o sector da grande distribuição sofreu grandes variações com a abertura de aproximadamente 80 hipermercados e o encerramento de 24 mil lojas do comércio tradicional (Santos, 2012).

À semelhança com outros países, em Portugal o sector da distribuição tem um enorme impacto económico e social em virtude do número de empresas e trabalhadores envolvidos. De acordo com Associação Portuguesa de Empresas de Distribuição, no ano de 2012, as empresas de distribuição em Portugal empregavam cerca de 86 700 colaboradores (Quadro 2.1) (APED, 2012).

De acordo com os dados do INE, em 2011 existiam em Portugal mais de 156 mil empresas a realizar comércio retalhista, as quais empregam 451,6 mil pessoas e dinamizam um volume de negócios de 45 855 milhões de euros, sendo que 31% do volume de negócios corresponde a produtos de alimentação, bebidas e tabaco (Guerreiro, 2014).

A análise do número de empregados e das empresas de comércio a retalho também permite perceber o peso e importância que este sector tem na economia e no emprego. No entanto (Quadro 2.1), são as empresas de pequena dimensão, onde, de um modo geral, se enquadram os estabelecimentos comerciais de proximidade, que se destacam em número e em quantidade de funcionários. Por outro lado, as empresas de média e grande dimensão, são menos e com maior número de empregados, o que pode ser um indicador que apesar do domínio das grandes superfícies, o pequeno comércio ainda mantém uma posição de destaque neste aspeto. No entanto, é importante reafirmar que também aqui o sector tem vindo a apresentar sinais de crise, dada a perda consecutiva de postos de trabalho.

Quadro 2.1: Empresas por Dimensão de Pessoas ao Serviço.

Adaptado de Guerreiro (2014, p.43)

Empresas por Dimensão de Pessoas ao Serviço				
Dimensão da Empresa	Número de Pessoas ao Serviço	Empresas	Pessoal ao Serviço	Rácio de pessoal ao serviço por empresa
Pequena	0 - 49	252 767	580 788	2,3
Média	50 - 249	950	86 749	91,3
Grande	250 ou mais	118	134 007	1135

Apesar destes números o sector do comércio a retalho alimentar está em crise (Guerreiro, 2014, p.42). No entanto, de acordo com o estudo Nielsen Market Track de Janeiro de 2017, o mercado do grande consumo encontra-se em evolução positiva, mostrando um crescimento de 3,9% comparativamente com o mesmo período de 2016. Na análise por canal de distribuição, o comércio tradicional apresenta um crescimento superior aos hipermercados (6,5% contra os 2,2%), podendo ser justificado pela aposta dos grupos de distribuição no formato de proximidade, como pequenos supermercados, lojas de conveniência, e outros pontos de venda de pequena dimensão.

2.4. Estabelecimentos de Proximidade

Desde 1976 têm sido desenvolvidos uma série de estudos que demonstram o enfraquecimento dos mercados locais e a expansão dos grandes estabelecimentos (Davies, 1976; Dawson, 2000; Maroney, 1976). Como tendencialmente, estas grandes lojas eram localizadas fora do centro das cidades e as pessoas se deslocavam para ir a estas, a economia local ficou afetada, uma vez que os pequenos retalhistas não tinham como concorrer com estas lojas, quer em termos de número de produtos comercializados quer nos preços praticados por eles, o que proporcionou o encerramento de muitas lojas de comércio local/tradicional (Sansone & Colamatteo, 2017). Porém, com o crescimento rápido destes retalhistas, o mercado começou a ficar saturado e houve necessidade de se

alterar a dinâmica de mercado, a favor dos pequenos estabelecimentos (Afonso, 2018; Serafim, 2014; Santos, 2012).

Atualmente, o conceito de vizinhança está a evoluir para novos padrões que se desviam de interpretações prévias da literatura. O que anteriormente era identificado sob um aspeto puramente dimensional, hoje é sinónimo de proximidade relacional com o consumidor (Sansone & Colamatteo, 2017). Com a criação dos hipermercados na periferia das cidades, as pessoas tinham de se deslocar a estes locais de propósito para fazerem as suas compras. Devido ao aumento da concorrência entre distribuidoras, estas começaram a apostar na abertura de estabelecimentos comerciais de proximidade nos locais de residência e/ou de trabalho, onde existem elevados fluxos de tráfego pedonal, para assim estarem mais presentes na vida dos consumidores (Guerreiro, 2014).

Dawson (1995, p.10) caracteriza o comércio de proximidade como sendo uma loja localizada numa área de grande concentração de clientes, em que os produtos consigam prever as necessidades imediatas dos consumidores ou que possam ser alterados com frequência, e como esteja localizada numa área conveniente para os consumidores de modo a que estes consigam visitá-la com facilidade e com frequência para gerar a fidelidade dos clientes. Ou seja, este tipo de comércio não está baseado no tipo de produtos que vende, mas na resposta dos consumidores ao formato da loja. Segundo Wells (2018), existem cinco tendências que vão revolucionar este tipo de estabelecimentos, nomeadamente a rapidez nas compras, a venda de produtos saudáveis, o proporcionar experiências aos consumidores (*e.g.*: realçar certos produtos ou uma área de estar de modo a que os clientes passem mais tempo na loja), a atração de novos consumidores (*e.g.*: *millenials*), através de uma oferta personalizada que poderá tornar as compras numa experiência e assim promover a fidelidade deste clientes, e a introdução da tecnologia para melhorar e simplificar a experiência.

De acordo com Zentes et al. (2007, p.122), este tipo de estabelecimentos permitem uma forte presença de determinada insígnia e proporcionam ao retalhista vantagens consideráveis, por exemplo:

- Uma insígnia de retalho existente fortalece o reconhecimento da marca e a diferenciação da concorrência, uma vez que com a crescente sobrecarga de informações ao consumidor, marcas estabelecidas e bem conhecidas recebem mais atenção do que marcas desconhecidas;

- Do ponto de vista do consumidor, marcas fortes de retalho simplificam o processo de compra porque já existe algum conhecimento sobre o retalhista e os compradores não precisam procurar informações adicionais. Um perfil de marca bem definido pode estabelecer uma posição de preferência que permita ao retalhista minimizar a concorrência de preços;
- Marcas fortes exercem o fenómeno de *halo-effect*, ou seja, uma atitude geral positiva em relação à marca principal influencia positivamente a perceção de todas as marcas secundárias;
- Marcas fortes não representam apenas benefícios funcionais, mas também podem servir como dispositivos simbólicos, isto é, estas representam diferentes valores e características, podendo projetar uma imagem diferente da marca original.

Segundo os estudos Nielsen Market Track de Junho de 2012 a Janeiro de 2017, em Portugal, no ano de 2012, o comércio tradicional apresentava valores negativos na ordem dos -9,6%, com o encerramento de muitas lojas que não conseguiam competir com os grandes retalhistas, enquanto os hipermercados e os grandes supermercados apresentavam valores na ordem dos 3,6% e 4,4%, respectivamente. Em 2014, todas as formas de comércio registaram crescimentos em relação ao período homólogo, os hipermercados registando o maior crescimento (cerca de 0,7% em relação a 2013) e o comércio tradicional, apesar de ainda apresentar valores negativos, apresentou um crescimento de 0,2%. Nos anos seguintes (Figura 2.4), foi-se assistindo ao crescimento do comércio tradicional e diminuição dos hipermercados, como referido anteriormente.

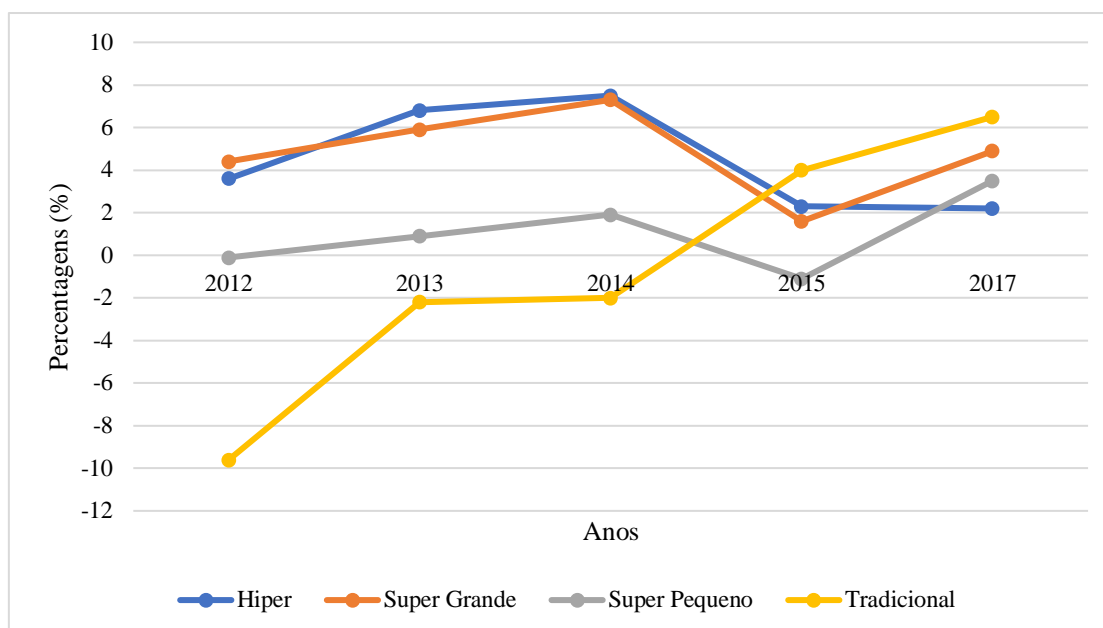


Figura 2.4: Evolução do crescimento dos sectores de comércio

Adaptado de Nielsen Market Track, 2017, 2015, 2014, 2013, 2012

De acordo com Paulo Proença (VASP, 2017), devido à crescente pressão do tempo sobre os consumidores, a tendência é de estes começarem a optar pelo comércio de proximidade em detrimento das grandes superfícies. Estas lojas, com o uso de novas tecnologias e o apoio das grandes distribuidoras vão permitir a disponibilização de novos serviços e produtos aos consumidores, com o benefício do atendimento personalizado e a maior proximidade aos clientes.

Atualmente, os estabelecimentos de proximidade podem apresentar dois formatos: as pequenas lojas de comércio tradicional (*e.g.*: mercearias de bairro), que ainda conseguem sobreviver apesar da pressão de outras lojas de cariz mais forte no mercado; e os formatos de *franchising* e cooperação comercial entre grupos empresariais e proprietários de lojas alimentares de pequena e média dimensão. Estas lojas “assentam na reconversão e adaptação das lojas de comércio tradicional ao *layout* e *naming*” de determinado grupo, replicando lojas, assim como à oferta de produtos desse grupo (Jorge, 2011), podendo ou não existir *royalties* ou taxas sobre as vendas realizadas (*e.g.*: lojas Meu Super).

Por *franchising* ou franquia entende-se “o sistema de colaboração entre empresas distintas e independentes ligadas por um contrato através do qual uma das empresas – o franqueador – concede a outra – o franqueado – o direito de explorar, em condições

preestabelecidas e bem determinadas, uma marca ou fórmula comercial sinalizada no exterior por uma insígnia específica e prestando ao franqueado determinados serviços regulares” (Rousseau, 2008, p.124). Ou seja, o franqueador fica encarregue de conceder a insígnia, marca e logótipo, assim como o transmitir a experiência acumulada sobre o negócio e a venda da sua coleção de produtos e serviços, e o franqueado fica incumbido da gestão da loja, adquire o saber-fazer do negócio, garante a assistência técnica e financeira permanente, que lhe dará maiores possibilidade de êxito, e beneficia ainda de preços e condições de compra mais favoráveis (Figura 2.5).

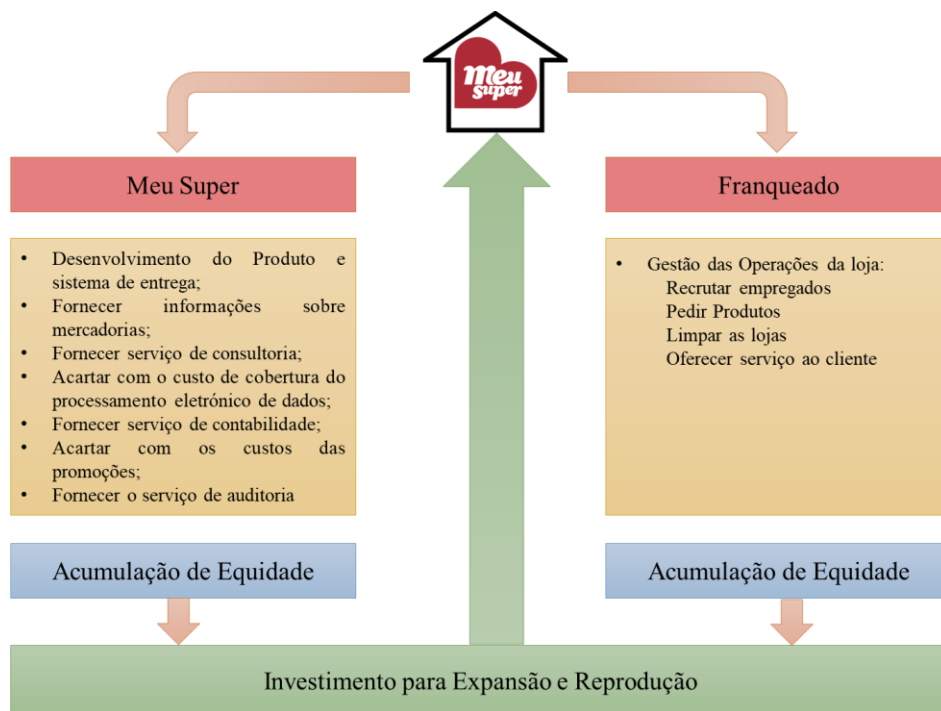


Figura 2.5: Funções da insígnia e do franqueado

Adaptado de Seven & i Holdings (2012, p.24)

Sendo o *franchising* um sistema de comercialização baseado numa colaboração entre empresas jurídicas, financeiramente independentes, este apresenta uma série de vantagens e desafios para ambas as partes.

No caso do franqueador, este tem como vantagens a divisão dos custos inerentes ao investimento necessário à abertura de novos pontos de venda, facilitando assim o crescimento em número de unidades e a expansão territorial da cadeia; o alargamento da sua influência geográfica, conquistando mercados sucessivamente mais vastos, e assim, minimizar a importância do fator distância, levando consequentemente ao fortalecimento

da insígnia; devido à partilha dos riscos financeiros entre ambas as empresas, este contrato vem tranquilizar o franqueador de que o franqueado apresente níveis de empenhamento no negócio substancialmente superiores àqueles que poderiam resultar caso fosse um simples trabalhador por contra de outrem; e ainda, com o *franchising*, a insígnia consegue controlar a concorrência, uma vez que o franqueador apenas poderá vender produtos e serviços comercializados, patenteados, licenciados ou autorizados pelo franqueador, não havendo guerras nas prateleiras. No entanto, ao contrário como acontece com o sucursalismo, as cadeias detentoras das marcas perdem parte do controlo do sistema, o que poderá possibilitar disputas e conflitos, entre as partes, ao nível das estratégias de gestão das unidades; e com a transferência do *know-how* do franqueador, este terá uma perda parcial de sigilo do negócio e, caso haja desistência por parte do franqueado, poderá transformar-se num veículo potencial de concorrência (Cachinho, 1999, p.244-247).

Ao nível do franqueado, este ganha um elevado sentimento de segurança, uma vez que não perde a independência e beneficia diretamente de estruturas, dos métodos e da notoriedade de redes comerciais importantes, permitindo gozar de um posicionamento privilegiado no mercado, correndo riscos substancialmente inferiores aos dos comerciantes independentes; recebe benefícios financeiros, quer pela rendibilidade imediata do negócio quer pelos efeitos favoráveis, pelo facto de estar integrado numa grande organização retalhista (e.g.: obter financiamento a taxa de juros menores nas instituições bancárias); e ainda, sendo o *franchising* um conceito profundamente testado nas suas distintas dimensões pelo franqueador, este contrato proporciona ganhos substanciais de tempo e de produtividade (e.g.: evitando possíveis erros que podem advir da falta de experiência, tanto a nível do negócio como da localização e organização do estabelecimento). Por outro lado, o contrato de *franchising* restringe fortemente a liberdade de ação do franqueado, restringindo os membros da rede no campo do desenvolvimento do *mix comercial*, das técnicas de *merchandising*, e da arquitetura e design do estabelecimento; a adesão a este contrato implica, por vezes, o pagamento de direitos de entrada e *royalties* periódicos; a participação neste sistema raramente permite aos membros da rede a transferência dos direitos adquiridos para terceiros, seja mediante a venda do estabelecimento ou do simples *trespasse* do negócio; e ainda, o franqueado será sempre visto como um trabalhador ao serviço da empresa franqueadora, independentemente do esforço efetuado para o sucesso do estabelecimento (Cachinho, 1999, p.247-250).

Este novo modelo de comércio permite aos comerciantes estarem mais próximos dos seus consumidores, trazendo benefícios aos clientes e aos comerciantes, uma vez que essa proximidade permite aos comerciantes perceber e antecipar as expectativas e necessidades dos consumidores, e melhorar/adequar a oferta a essas necessidades, tornando a oferta mais eficiente.

3. Geografia, SIG e *Geomarketing*

A geografia é uma ciência interdisciplinar, visto que compreende as ciências sociais e humanas e as ciências naturais, apresentando várias vertentes, sendo as principais a vertente física e a vertente humana. Devido a esta versatilidade esta ciência já se encontra bastante inserida no mundo empresarial, uma vez que a Geografia consegue perceber o território e os fenómenos que nele acontecem que podem condicionar o sucesso ou fracasso de uma empresa.

A localização de uma empresa condiciona muito o sucesso da mesma. Como referido anteriormente (subcapítulo 2.2), desde o final do século XIX, houve por parte de economistas o interesse sobre a localização dos serviços, provando que a localização é o fator-chave para a implementação de uma empresa. As teorias de localização e as manifestações económicas ocorrem invariavelmente do tempo e do lugar, em espaços caracterizados por fenómenos de aglomeração demográfica e centralidade espacial. Assim, a ciência locativa interessa-se em estudar e determinar a posição ótima dos agentes económicos no ambiente urbano. Com o aumento do número de cadeias de retalho, esta ciência torna-se essencial para observar a realidade do mercado e como este se comporta com a introdução de novos agentes. A ciência locativa deve incluir várias disciplinas, *e.g.*, a geografia, o marketing e a cartografia (= *Geomarketing*).

Este capítulo está reservado ao surgimento e importância do *Geomarketing*. Vários autores (Aranha & Figoli, 2001; Catunda, 2010; Latour & Floc'h, 2001) concordam que a introdução da componente espacial nas teorias económicas foi o fator impulsionador para o aparecimento do *Geomarketing*.

Cavion e Philips (2006) afirmam que o surgimento do *Geomarketing* se deveu a quatro grandes acontecimentos: o desenvolvimento de teorias económicas, o aparecimento do marketing e a sua colaboração com a Geografia, a evolução dos TIC e a criação dos SIG (Figura 3.1). Estes destacam ainda que a geografia, a cartografia e o marketing estão diretamente interligados de modo a que cada área contribua para um resultado mais eficiente dentro do *Geomarketing*.

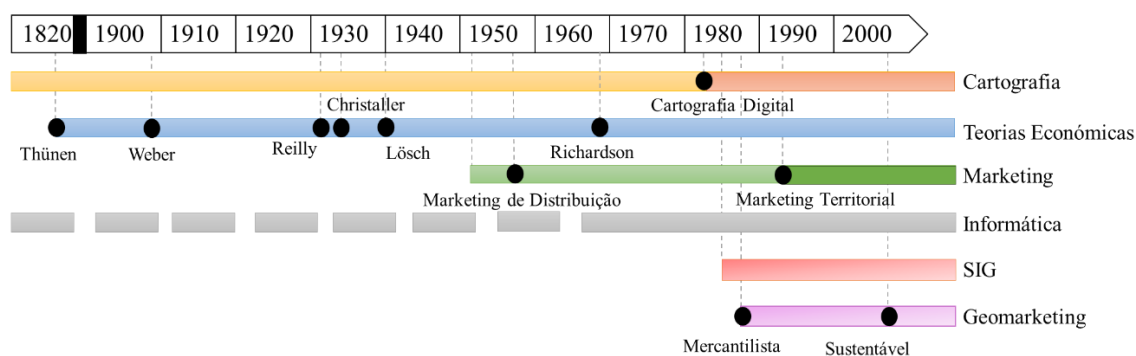


Figura 3.1: Ordem cronológica do surgimento do *Geomarketing*, por ramos de conhecimento

Fonte: Cavion e Philips (2006, p.3)

3.1. *Geomarketing*

O marketing, como referido anteriormente, era definido como a arte e a ciência de escolher os mercados-alvo e de conquistar, reter e cultivar clientes. À medida que a sociedade evoluiu, o comportamento dos clientes vai-se alterando, levando a que esta ciência se adaptasse e procurasse novos conhecimentos para aumentar a sua eficiência nas decisões empresariais, para isso, foram incorporando novas técnicas, conceitos, ferramentas e cooperando com outras áreas de estudo, entre elas a estatística, psicologia, sistemas de informação e a geografia (Silva & Coelho, 2003, citados por Bernardes, 2017).

Com esta necessidade do estudo dos clientes e das localizações surge o *Geomarketing*. Na década de 1980, o *Geomarketing* surge como a confluência entre o Marketing e a Geografia, adicionando ao marketing as componentes físicas e humanas que a Geografia proporciona. No entanto, este conceito é mais amplo, englobando outros elementos e ciências como a informática, a estatística e a cartografia (Yrigoyen, 2003). De acordo com Cliquet (2006), o *Geomarketing* necessita de assumir funções de multidisciplinaridade, entre elas, a cartografia, a geografia e o marketing; que conseguem interligar visões distintas para a resolução de um mesmo problema (Figura 3.2).

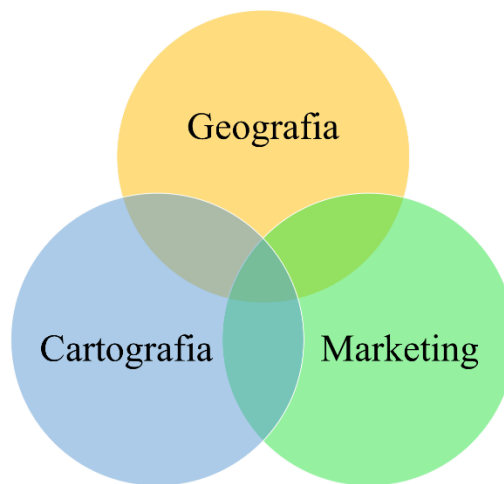


Figura 3.2: Multidisciplinaridade do *Geomarketing*

Assim, foram vários autores que teceram as suas considerações relativamente à definição de *Geomarketing*. Latour e Floc'h (2001) definiram *Geomarketing* como um método desenvolvido para interpretar dados gerados de programas de computador, métodos estatísticos e gráficos, com o objetivo de desenvolver informações pertinentes à tomada de decisão (gráficos, tabelas e cartografia digital). Por outro lado, Yrigoyen (2003), salienta ainda que o *Geomarketing* permite conceber novas estratégias empresariais, uma vez que a distância exerce influência sobre as atividades económicas. Para Cliquet (2006), os estudos em *Geomarketing* estão assentes em três grandes blocos: técnicas de codificação, delimitação geográfica e funções estatísticas, que em conjunto permitem manipular os dados de forma a compreender e planejar o território.

A origem deste termo é relativamente recente, sendo datado da década de 1980, quando se iniciaram as investigações nesta temática. No entanto, as bases do *Geomarketing*, remontam ao final do século XIX e princípios do século XX, com os estudos da economia espacial que eram fundamentados na hipótese de a distância geográfica exercer influência sobre a atividade económica. No final dos anos 1980, inícios dos anos 1990 houve uma grande procura por este tipo de análises espaciais, de acordo com Yrigoyen (2003, p.8) essa procura deveu-se:

- 1) Ao *boom* no campo das ciências sociais e nas teorias económicas, dando origem à “nova geografia económica”. Este interesse na questão espacial não se limitou à economia, dispersando-se para outras ciências sociais;
- 2) À crescente disponibilidade de grandes bases de dados socioeconómicos de natureza territorial, georreferenciadas e, muitas vezes, gratuitas;

- 3) Ao desenvolvimento de uma tecnologia eficiente e de baixo custo, capaz de manipular observações georreferenciadas: Sistemas de Informação Geográfica (SIG), e em geral, *software* para análise de dados espaciais.

Foi o desenvolvimento das Tecnologias de Informação e Comunicação (TIC) que despontaram o rápido crescimento e utilização do *Geomarketing*. A utilização dos SIG possibilitou o armazenamento eficiente e uma rápida recuperação e visualização interativa de conjuntos de dados geográficos, oferecendo uma boa solução para as empresas produzirem informações relevantes e úteis para as tomadas de decisão.

3.2. Sistemas de Informação Geográfica (SIG)

A primeira aplicação dos atualmente conhecidos como sistemas de informação geográfica remonta a 1884 com o estudo, por parte do físico Inglês John Snow, da epidemia da cólera em Londres. Através de um mapa com a localização da população afetada e dos possíveis modos de transmissão da doença, John Snow, descobriu que a causa dos óbitos por cólera partia de certos poços de água contaminados, localizados na cidade.

Com o aparecimento dos primeiros computadores, no início da década de 60, apareceram também os primeiros SIG. O primeiro SIG foi criado por Roger Tomlinson e desenvolvido no Canadá, em 1966, com o intuito de analisar, processar e criar o inventário dos dados geográficos relativos ao solo, a nível nacional. Este pretendia reduzir os custos e a falta de eficiência associados ao tratamento de grandes volumes de dados (Teixeira, 1993). Ao longo das décadas seguintes, os SIG evoluíram tanto a nível de *software* como a nível de inovações, quer nas aplicações dos SIG quer na participação dos utilizadores que passaram a ter uma participação ativa no desenvolvimento de *softwares* e na aplicação dos SIG em diversas áreas de interesse e sectores de atividade. Esta tecnologia tem sido utilizada por diversas organizações quer públicas quer privadas, revelando-se como uma ferramenta importante no planeamento urbano, saúde pública, segurança, ecologia, redes de transportes, demografia, gestão de recursos e em muitas outras organizações.

Desde o seu aparecimento, os Sistemas de Informação Geográfica (SIG) têm sido alvo de várias definições, umas mais genéricas e outras mais específicas como observável no Quadro 3.1.

Quadro 3.1: Principais definições de SIG

Autor	Definição de SIG
(Deuker, 1979)	Um tipo de sistema de informação, em que a base de dados é formada por características, atividades ou eventos distribuídos espacialmente
(Burrough, 1986)	Um poderoso conjunto de ferramentas para aquisição, armazenamento, recuperação e exibição de dados do mundo real para determinados propósitos
(Cowen, 1988)	Um sistema de apoio à decisão que envolve a integração de dados espacialmente referenciados, num ambiente para resolução de problemas
(Parker, 1988)	Uma tecnologia de informação que armazena, analisa e exibe dados espaciais ou não - SIG é de facto uma tecnologia e necessariamente não é limitada a um simples e bem definido sistema de computador
Federal Interagency Coordinating Comitee, 1988	Um SIG é um sistema computacional desenvolvido para permitir que os utilizadores recolham, administrem e analisem grandes volumes de dados referenciados, de modo a resolver problemas complexos de administração e planeamento
(Aronoff, 1989)	Qualquer conjunto de procedimentos manuais ou baseados em computador destinados a armazenar e manipular dados georreferenciados
Centro Nacional de Informação Geográfica, 1990	Conjunto de dados localizados no espaço, estruturados de modo a extrair informações relevantes para a tomada de decisão
(Worboys, 1995)	Um sistema de informação baseado em computador que permite a captura, modelação, manipulação, recuperação, análise e apresentação de dados georreferenciados

Apesar das diversas definições, não existe uma que seja universal pois as definições dependem da forma como são desenvolvidas (como no caso da definição de Aronoff (1989)) e do propósito para que foram criadas (como a de Cowen (1988)), isto é, cada autor, suprarreferido, adequou a sua definição ao trabalho que estava a desenvolver, o que não invalida a sua definição.

Os SIG são um sistema de informação, ou seja, são caracterizados por dois componentes: um conjunto de dados e um conjunto de técnicas, procedimentos e regras capazes de estruturar, relacionar e interpretar esses dados, com o intuito de transformá-los em informação (Teixeira, 1993). Como são sistemas de informação geográfica essa informação apresenta a componente espacial. Em 2005, Longley et al., definiram os componentes de um sistema de informação geográfica, nomeadamente, *hardware*

(computadores, workstations, etc.), módulos de *software* (ArcGIS, MapInfo, QGIS, etc.), dados (espaciais e alfanuméricos), procedimentos e as pessoas (as que trabalham com os SIG e os utilizadores).

Os SIG são a ferramenta mais utilizada para analisar, visualizar e mapear dados espaciais, em grande parte devido à facilidade dos SIG para usar e combinar diversas bases de dados, transformando-os numa única base de dados (Pettit & Pullar, 1999), permitindo assim que os utilizadores consigam sobrepor informação de diversas fontes com elevada precisão. De acordo com Church (2002), os sistemas de informação geográfica têm sido muito úteis em abordagens analíticas espaciais e na formação de uma interface com o campo da ciência da localização. Vários estudos fornecem uma visão geral dos impactos dos SIG em trabalhos no campo da localização, com a aplicação de modelos, desenvolvimento e criação de vários métodos que podem ser usados na modelação do uso do solo (Malczewski, 2004).

A crescente preocupação com o planeamento de negócios e serviços, de modo a atrair consumidores, juntamente com os avanços nos SIG e técnicas de análise espacial, levou à elevada utilização destas técnicas na área de planeamento e gestão de negócios (Longley & Clarke, 1995). A utilização de sistemas de informação geográfica e *softwares* de análise espacial pode facilitar a resolução de problemas em vários temas, *e.g.*, análise de localização do retalho, marketing localizado, etc. (Niti, 2007).

Sendo os sistemas de informação geográfica fundamentados nos princípios da Geografia, estes conseguem responder a questões como: “Onde está ...”, “Qual a melhor localização para...?”, “Qual o melhor caminho para...?” (Rhind, 1990 citado por Teixeira, 1993, p.18). Assim, a informação geográfica assume-se de grande importância para conhecer e compreender a realidade, o que pode explicar o facto que se estima que 80% da informação pode ser georreferenciada (Worrall, 1991). Esta constatação, segundo Boudot (1999), esteve na origem do *Geomarketing*.

3.3. SIG, *Geomarketing* e o comércio a retalho

De acordo com Fagudes et al. (2009) citado por Rodeski (2010), o *Geomarketing* é sustentado por dois pilares, nomeadamente, os Sistemas de Informação Geográfica (SIG) e o Marketing, e as suas principais aplicações assentam no suporte técnico para interpretar questões mercantilistas de maior significância. Com a introdução dos SIG no

processo de tomada de decisão, a componente espacial revelou-se uma variável descritiva e explicativa, uma vez que, segundo Harris et al. (2005), as pessoas que compartilham um ambiente geográfico tendem a partilhar comportamentos, atitudes e hábitos de consumo, tornando o estudo da localização um fator determinante na análise da população.

Para fazer um estudo de *Geomarketing* são necessários os seguintes elementos: bases de dados, informação cartográfica e os SIG para o processamento e gestão da informação. Como se pode verificar na Figura 3.3, os elementos constituintes do *Geomarketing* são a informação estatística e cartográfica, o tratamento da informação e os estudos de mercado.

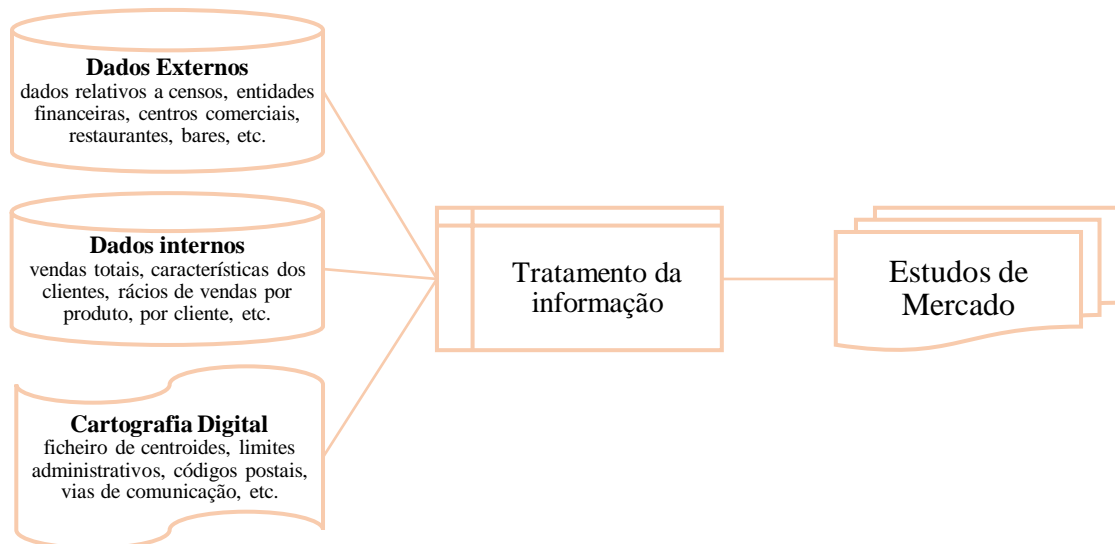


Figura 3.3: Elementos do *Geomarketing*

Fonte: Yrigoyen (2003, p.8)

A informação estatística, que pode ser interna das empresas (vendas, transações, clientes, etc.) ou externa (dados estatísticos, censos, etc.), constitui a base de todos os estudos de mercado. Juntamente com a informação estatística, a cartografia digital é fundamental para este tipo de estudos, uma vez que permite visualizar e tratar estatisticamente dados de diferentes zonas e a diferentes escalas ou âmbitos geográficos. Deste modo, os SIG aparecem como a tecnologia que faz com que seja possível a visualização interativa, exploração, armazenamento e recuperação rápida dos dados geográficos, exibindo-os sobre formas de mapas, tabelas, gráficos e textos. No entanto, é essencial que se conheçam as relações entre os diferentes tipos de dados, para isso é

necessário tratar a informação, especificamente para observar as relações entre os dados terá de se recorrer a uma análise estatística exploratória espacial que se aplica a grandes volumes de dados que é denominado de *Data Mining* (Yrigoyen, 2003). Com esta análise e a modelação espacial da informação será possível obter resultados que permitirão fazer os mais variados estudos de mercado, *e.g.*, o estudo da localização dos clientes, áreas de influência e a deteção de lacunas de mercado.

O poder dos SIG reside na sua capacidade de integrar informações espacialmente relacionadas, de manipular esta informação em relação a todos os atributos e locais, de realizar análises espaciais simples e para preparar facilmente mapas e gráficos apelativos e informativos que permitem exibir e compreender informações relacionadas com o espaço. A correta utilização das ferramentas dos SIG pode auxiliar na resolução dos problemas de localização, devido à sua capacidade para integrar grandes volumes de informações espaciais de diversas origens. Deste modo, consegue-se analisar possíveis localizações com base no tempo de condução, variáveis demográficas e socioeconómicas, informações competitivas e informações do cliente (Mendes & Themido, 2004), tornando-se uma ferramenta importante no suporte à tomada de decisão.

Na atualidade, a escolha de uma localização para um estabelecimento comercial é mais desafiadora, uma vez que o aumento da concorrência limita os espaços disponíveis. Cheng et al. (2007) argumentaram que os sistemas de informação geográfica se tornarão ferramentas essenciais para empresas de retalho para desenvolver sistemas de decisão e seleção de locais para novos estabelecimentos. De acordo com Clarke (1998), o uso dos SIG nas teorias de localização do retalho tornou-se mais generalizado devido à importância das características geodemográficas, uma vez que os SIG tornaram possível a segmentação dos consumidores e, consequentemente, do mercado de acordo com as características geodemográficas da população.

3.4. Geodemografia

De uma forma muito generalista Sleight (1997, p.16) define geodemografia como a “análise das pessoas segundo o sítio onde vivem”. De acordo com Johnson (1989), citado por Baviera-Puig et al. (2016), as características geodemográficas classificam a população consoante o local onde vivem em vez de consoante os critérios socioeconómicos, como a classe social ou o rendimento disponível. Ambos os conceitos

partem do princípio que as pessoas tendem a agrupar-se a pessoas similares, por exemplo, em termos de etnicidade ou estatuto social, o que permite tecer considerações sobre o consumo afeto a essa população.

Estes conceitos sugerem que o local onde estão inseridos ditam quem são e que fornecem informações sobre como as pessoas vivem. É uma ideia simples, que se mostrou de valor comercial e catalisador de uma indústria em rápida expansão e globalização (Harris et al., 2005). Autores como Weber (1985), Curry (1993) e Sleight (1997), citados por Harris et al. (2005), reconheceram a Geodemografia como uma ferramenta importante nos negócios, uma vez que esta oferece um tipo de análise fácil e funcional num ambiente de tomada de decisão.

Leventhal (1993, p.223), reconhecia o potencial da geodemografia em fornecer informação para estratégias de marketing, planeamento e comunicação, apresentando exemplos da sua aplicação em pesquisas de mercado em três sectores importantes do Marketing:

- 1- Projeto de pesquisa: aproveitamento da seleção e estratificação da amostra, realizada através da Geodemografia, para realizar inquéritos a grande escala;
- 2- Planeamento de retalho: através do conhecimento dos tipos de pessoas que vivem em determinada área de influência, permite o conhecimento do desempenho da loja e pode auxiliar na decisão da localização da loja;
- 3- Marketing direto: a seleção de clientes potenciais pode ser melhorada através da geodemografia, seja através de correio, distribuição porta-a-porta ou chamadas telefónicas (*sales calls*).

Estes três temas correspondem à pesquisa de mercado, análise de mercado e fluxos de marketing direto identificados por Curry (1993, p.200) e, ainda, um quarto papel da geodemografia na análise de publicidade e dos media, que se encontram resumidos na Figura 3.4.

A Figura 3.4 oferece uma perspetiva do marketing em aplicações geodemográficas. É de salientar que as análises de vizinhança, muito utilizadas em Geodemografia, não são exclusivas para uso comercial. De modo a analisar as necessidades de análises de mercado e perfis de consumo, estas deverão ser utilizadas também no sector público (examinar padrões espaciais de crime, saúde, etc.) e no

planeamento (projetar serviços para antecipar/combater problemas sociais). No entanto, este relatório centra-se no comércio por isso abordar-se-á mais esta vertente.



Figura 3.4: Algumas aplicações comerciais da geodemografia

Adaptação de Harris et al. (2005, p.5)

O termo geodemografia deriva da junção entre a demografia e as dimensões geográficas da população. Este conceito é de elevada importância em atividades que comprometem grandes quantidades de capital, uma vez que nem toda a população que está numa determinada área será um cliente de uma determinada empresa. Torna-se, assim, essencial conhecer os possíveis clientes e como estes estão distribuídos para escolher a melhor localização possível para a empresa.

Kim e Choi (2013) afirmam que a localização afeta a experiência comercial dos clientes, o que vai influenciar a sua fidelidade. Esta afirmação é especialmente característica do comércio a retalho. De modo a rentabilizar o comércio a retalho é essencial ter na sua área de influência o maior número de grupos de indivíduos que tenham hábitos de consumo similares. Deste modo, com um *target market* definido, o retalhista pode ter em *stock* os produtos necessários e providenciar serviços específicos para atender às necessidades desses potenciais clientes.

A utilização de sistemas geodemográficos são especialmente úteis para os retalhistas que têm os seus clientes concentrados em certos segmentos geodemográficos,

e são essenciais para encontrar localizações certas para a sua instalação. Este tipo de informação também permite aos retalhistas monitorizar variações no padrão da procura entre diferentes grupos geodemográficos (Murad, 2015).

A definição ampla de geodemografia (uso de variáveis demográficas dentro de áreas geográficas) implica que se pode usar informações censitárias para fazer análises geodemográficas. O recenseamento da população fornece um registo fiável das características socioeconómicas da população, assim como a sua distribuição numa determinada área, conseguindo, deste modo, apoiar análises geodemográficas (Harris & Longley, 2000). Através do uso destas variáveis é possível realizar-se a segmentação geodemográfica da população.

A segmentação geodemográfica refere-se à classificação dos consumidores consoante o tipo de área residencial onde vivem. Baseia-se, portanto, na diferenciação de áreas residenciais de acordo com as características demográficas, socioeconómicas ou mesmo psicográficas de seus residentes. A segmentação geodemográfica está sustentada em dois princípios subjacentes: o primeiro, em que áreas residenciais similares apresentam hábitos de consumo semelhantes e, consequentemente, respostas similares aos estímulos do marketing; e o segundo, em que indivíduos com características semelhantes tendem a residir nas mesmas áreas e compartilhar os mesmos ambientes, ou seja, as áreas residenciais tendem a ser internamente homogéneas para que seus residentes não difiram significativamente de um perfil (González-Benito & González-Benito, 2005).

A importância da caracterização geodemográfica dos clientes dos estabelecimentos de retalho tem sido repetidamente referida na literatura (Birkin et al., 2002; Clarke, 1998; González-Benito, 2002; Harris et al., 2005; Murad, 2015; O'Malley et al., 1997; Zentes et al., 2011), sendo que muitos autores enfatizaram a importância da geodemografia, a seleção de mercados geográficos e a localização das lojas de retalho como aplicações especialmente relevantes. A segmentação geodemográfica permite aos retalhistas medir a heterogeneidade espacial do mercado, distinguindo as zonas mais atraentes para cada tipo de loja.

Deste modo, a segmentação geodemográfica apresenta-se como uma boa ferramenta analítica para a escolha de possíveis localizações para lojas de retalho, resultado da sua capacidade de diferenciar mercados geográficos de acordo com o tipo de consumidores, e assim, fazer um ajuste entre os consumidores e a loja projetada.

Uma vez que o foco da segmentação geodemográfica é a classificação dos consumidores de acordo com a sua residência, as informações fornecidas por esses esquemas de classificação dizem respeito à avaliação do local da loja em relação à localização relativa aos consumidores. No entanto, estas características não só auxiliam na seleção de uma localização para um estabelecimento comercial, como influenciam a implantação de instalações complementares localizadas numa área de mercado e funcionam como indicadores do tipo de concorrência que aí se encontra.

A escolha da localização para um estabelecimento com base nas características das áreas de mercado deriva necessariamente de uma correlação significativa entre perfis geodemográficos, formatos e cadeias de retalho. Portanto, a segmentação geodemográfica pode facilitar a avaliação do local de retalho em relação à localização relativa aos concorrentes (Murad, 2015).

3.5. Decisão de localização do retalho

A competitividade entre empresas no comércio a retalho tornou-se tão exigente que algumas empresas precisam de múltiplas localizações para penetrar no mercado (Cardoso, 2011). De modo a possibilitar essa penetração surgiram os modelos de localização que podem permitir às empresas alguma vantagem perante a concorrência. Os modelos de localização já são utilizados como ferramentas de auxílio à decisão espacial há alguns séculos, e têm sido alvo de estudos e aprimoramentos ao longo dos anos. Estes modelos procuram localizar equipamentos, escolhendo, de um conjunto de locais possíveis, o melhor local para um equipamento, tendo em consideração os clientes que devem ser servidos, otimizando um determinado critério: maximizar a utilização ou minimizar as distâncias (Cabral, 2012; Costa, 2010; Lorena et al., 2001). De acordo com Cliquet (2006, p.149), os modelos de localização, principalmente os localização-alocação, contêm os seguintes componentes:

- Uma função objetiva para ser otimizada consoante os locais disponíveis;
- Áreas de procura reduzida a um ponto central onde a procura de bens e serviços está concentrada;
- As possíveis localizações, a acessibilidade e a infraestrutura;
- Distância ou tempo de acesso;

- A regra da distribuição, ou o modo de como os clientes escolhem entre as várias opções de oferta.

O sector do retalho é caracterizado por ter um ambiente dinâmico e incerto, o que leva a que as empresas retalhistas necessitem de antecipar e planear o mercado, e de se impor à concorrência. A natureza competitiva do comércio retalhista e as variadas técnicas empregues pelos retalhistas para encontrar as melhores localizações, fizeram com que os sistemas de informação geográfica começassem a fazer parte da tomada de decisões (Davies & Clark, 1994). No complexo processo de decisão que envolve uma grande variedade e volume de informação, métodos de visualização são muito úteis, assim os SIG tornam-se uma ferramenta essencial para o desenvolvimento do processo de decisão para a localização de estabelecimentos (Mendes & Themido, 2004). Os SIG não são usados apenas para estudos de localização e captação, sendo utilizados para outros sectores no retalho como a gestão de produtos, *merchandising*, comunicação de marketing (publicidade) e marketing relacional (O'Malley et al., 1997). Devido à utilização dos SIG, desde 1960, as metodologias empregues para a localização do retalho tornaram-se mais aprimoradas (Birkin et al., 2002), uma vez que os SIG contribuíram muito para a eficácia e precisão do planeamento de retalho e do marketing (Niti, 2007).

A seleção de locais para estabelecimentos de comércio a retalho é uma das decisões mais importantes na venda a retalho, uma vez que uma boa localização pode ser um elemento-chave para atrair clientes a determinado ponto de venda. No entanto, as decisões de localização destes estabelecimentos são altamente complexas devido ao grande número de fatores que devem ser considerados, assim como os custos associados à localização. Na Figura 3.5 encontram-se alguns dos fatores que podem influenciar o sucesso de uma empresa de retalho alimentar.

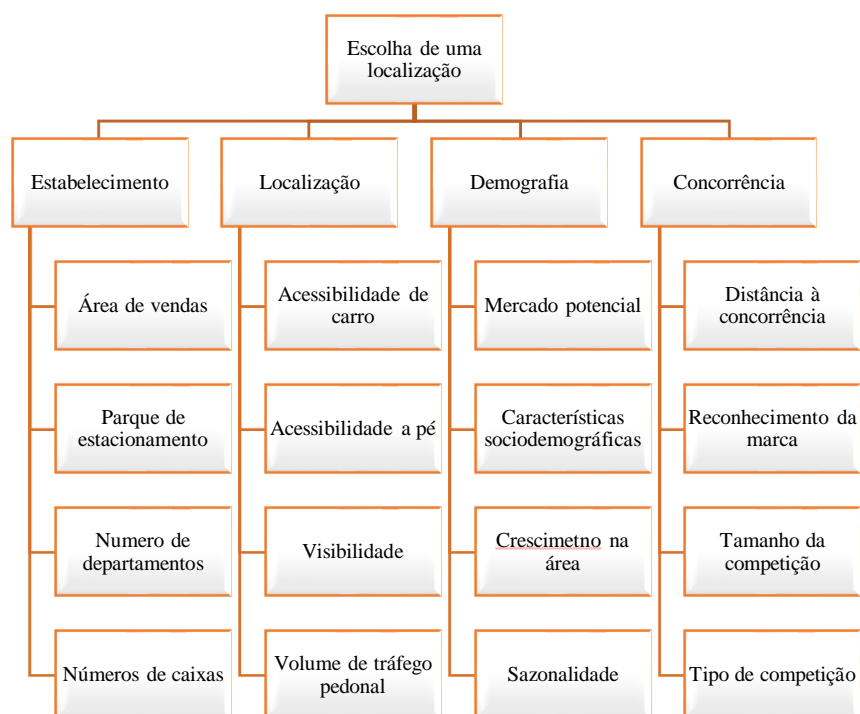


Figura 3.5: Fatores que afetam o sucesso de um estabelecimento de comércio de retalho alimentar

Fonte: Roig-Tierno, et al. (2013, p.194)

Estes fatores vêm reforçar a ideia que a seleção de um local para um estabelecimento é uma decisão que carece de tempo e capital a longo prazo, devido à sua natureza fixa, não podendo ser alterada em curto espaço de tempo como os restantes elementos do *marketing mix* (Zentes et al., 2011).

Quando se aborda estratégias de localização de estabelecimentos, pensa-se na abertura de novas lojas, no entanto as decisões de localização são mais abrangentes, uma vez que se referem a toda a estrutura física dos pontos de vendas a retalho, podendo, entre outras, decidir-se em relação à extensão do espaço das lojas existentes; à deslocalização ou ao movimento de uma loja de um lugar para outro; ou ao encerramento de lojas individuais (Hernández & Bennison, 2000, citados por Zentes et al., 2011).

A abertura de novas lojas compreende o tipo de decisão mais complexa, pois, geralmente, é o ponto de partida das atividades numa área geográfica específica. Nesta dissertação, pretende-se abrir novas lojas da insígnia Meu Super (MS) e, por isso, com recurso a ferramentas SIG, irá avaliar-se as diferentes localizações disponíveis com base em funções analíticas, incorporando previsões, objetivos e benefícios de cada uma delas.

4. Informação de suporte à análise

Como em qualquer estudo de *Geomarketing* são necessárias bases de dados, informação cartográfica e os SIG para o processar e gerir a informação. Assim sendo recolheram-se os dados presentes no Quadro 4.1 de modo a conseguir atingir os objetivos pretendidos.

Quadro 4.1: Dados utilizados

Dados	Tipo de elementos	Fonte de informação a utilizar
Rede Viária	Linhas com os diferentes tipos de estradas	Nokia 1:10 000
Lojas Meu Super	Pontos com a localização das lojas Meu Super	Websites das diferentes empresas, Google Earth, 1: 10 000
Lojas Pingo Doce	Pontos com a localização das Pingo Doce	
Lojas Lidl	Pontos com a localização das lojas Lidl	
Lojas Minipreço	Pontos com a localização das lojas Minipreço	
Lojas Amanhecer	Pontos com a localização das lojas Amanhecer	
Lojas Meu Super encerradas	Pontos com a localização das lojas Meu Super encerradas	
BGRI - Censos da população	Polígonos com dados sobre a população	INE 2011 1:10 000
Carta de Ocupação do Solo 2010	Polígonos com o tipo de ocupação do solo	DGT 2010 1:25 000

A aquisição dos dados referentes às lojas, quer as que se pretende expandir, nomeadamente as lojas ‘Meu Super’, como as que foram consideradas as suas competidoras diretas, nomeadamente o Pingo Doce, Lidl, Minipreço e Amanhecer, respetivamente, consistiu na criação de um inventário com informações referentes às moradas, áreas de loja, códigos postais, coordenadas geográficas, estado da loja (encerrada, em funcionamento e propostas), nome e número de identificação da loja, data de abertura e o grupo e cadeia da loja, de cada loja. Para as lojas Meu Super foi, ainda, possível obter dados referentes às vendas brutas e fidelizadas, número de clientes fidelizados e número de insígnias concorrentes nas respetivas áreas de influência. Para este efeito, foram recolhidas as lojas que se encontravam abertas ou que tinham encerrado até ao ano de 2015 para que não houvesse discrepância nos dados, visto que os dados adicionais referentes às lojas Meu Super correspondiam ao ano de 2015.

Com o inventário das lojas feito, tomou-se nota de duas situações. Para alguns casos foi possível obter coordenadas geográficas, no entanto noutros casos apenas foi possível obter o seu endereço. Para os anteriores, teve de se recorrer a uma técnica denominada de “Geocodificação” (Figura 4.1). A Geocodificação é um processo, de georreferenciação (Goldberg et al., 2007), que permite obter um código da localização específica (coordenadas) da informação a tratar (*e.g.* moradas). Ou seja, a Geocodificação permite associar uma morada ou nome de um local a dados espaciais (Polido, 2014), permitindo, através deste processo, encontrar padrões de análise (Rocha & Henriques, 2014).

A Geocodificação pode ser de diversos tipos, que terão diferentes níveis de precisão e/ou exatidão. Para este projeto o método utilizado foi a Geocodificação direta aproximada, visto que esta corresponde à determinação do centro geométrico (centroide) de um elemento de referência para obter o ponto de localização (Rocha & Henriques, 2014), técnica utilizada para gerar os pontos com a localização das lojas (a expandir e da concorrência) em Portugal Continental.

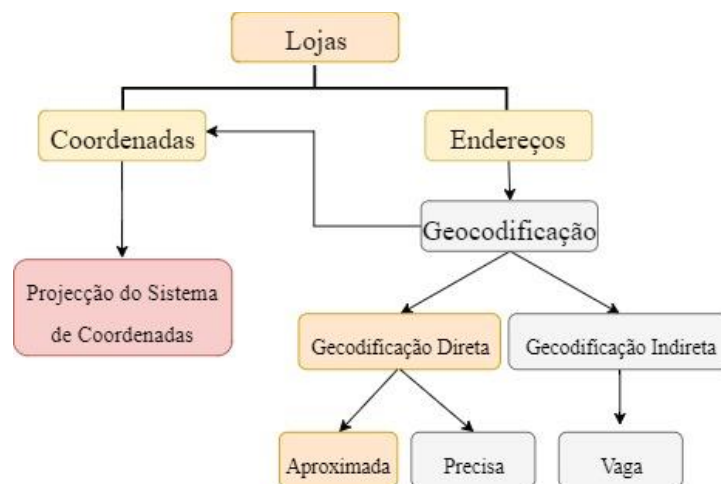


Figura 4.1: Esquema do processo de georreferenciação e métodos de Geocodificação

Com as coordenadas geográfica das lojas e a sua localização no mapa, foi necessário uniformizar o sistema de coordenadas, uma vez que os diferentes tipos de informação apresentavam sistemas de coordenadas diferentes. Optou-se por utilizar o sistema de projeção PT-TM06/ETRS89 (Portugal – *Transverse Mercator* 2006/*European Terrestrial Reference System* 1989), devido a este ser o sistema de coordenadas oficial português desde 2006 por diretiva da União Europeia.

Em relação às restantes variáveis, com os dados demográficos provenientes dos censos 2011 realizados pelo Instituto Nacional de Estatística, foi possível obter dados que permitiram criação de um perfil cliente. Estes juntamente com a carta de ocupação do solo proveniente da Direção Geral do Território, possibilitaram a realização da cartografia dasimétrica permitindo uma melhor qualidade na caracterização do território. E por último, a rede viária, proveniente da Nokia, e posterior criação da *Network Dataset*, permitiram a criação de mais variáveis e a realização de modelos localização-alocação.

Com os dados recolhidos foi possível criar as variáveis necessárias para a concretização deste projeto. Nos subcapítulos seguintes serão aprofundadas as variáveis e algumas metodologias empregues para a criação das mesmas. No capítulo seguinte (*Location Analytics*) serão abordadas as restantes metodologias e efetuados modelos preditivos. Na Figura 4.2, pode observar-se o esquema processual que será a base do trabalho realizado e que se irá abordar pormenorizadamente nos próximos capítulos.

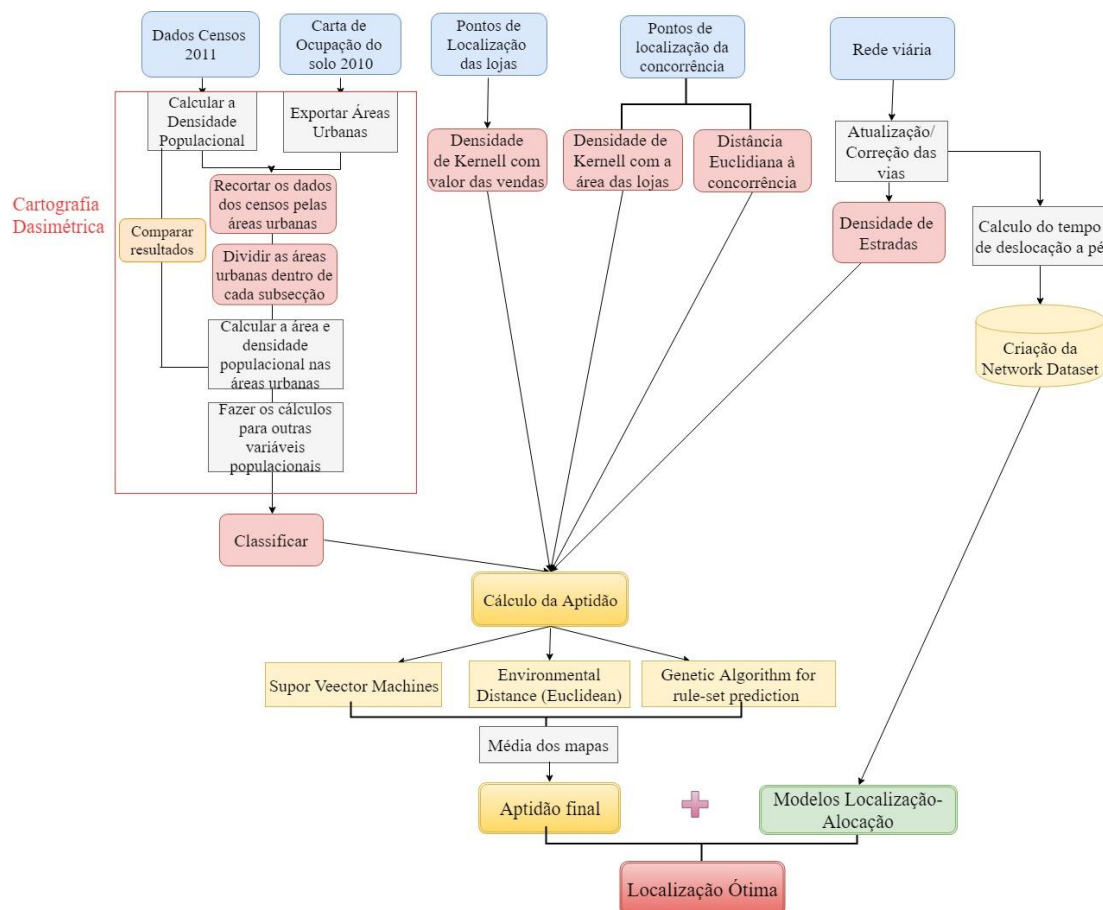


Figura 4.2: Esquema processual

4.1. Caracterização da Oferta

Em Portugal Continental existem inúmeras insígnias, no que se refere ao comércio de retalho alimentar. No entanto, algumas têm mais visibilidade e presença do que outras e por isso teve-se especial atenção à escolha da concorrência. Foram recolhidas 32 insígnias que poderiam ser concorrentes às lojas Meu Super. No entanto, visto que as lojas em estudo nunca apresentam uma área superior a 540 m², nem todas poderiam ser consideradas concorrentes diretas. Não obstante, na Figura 4.3 podem observar-se as cinco cadeias de supermercado com mais representação em Portugal Continental, no ano de 2015.

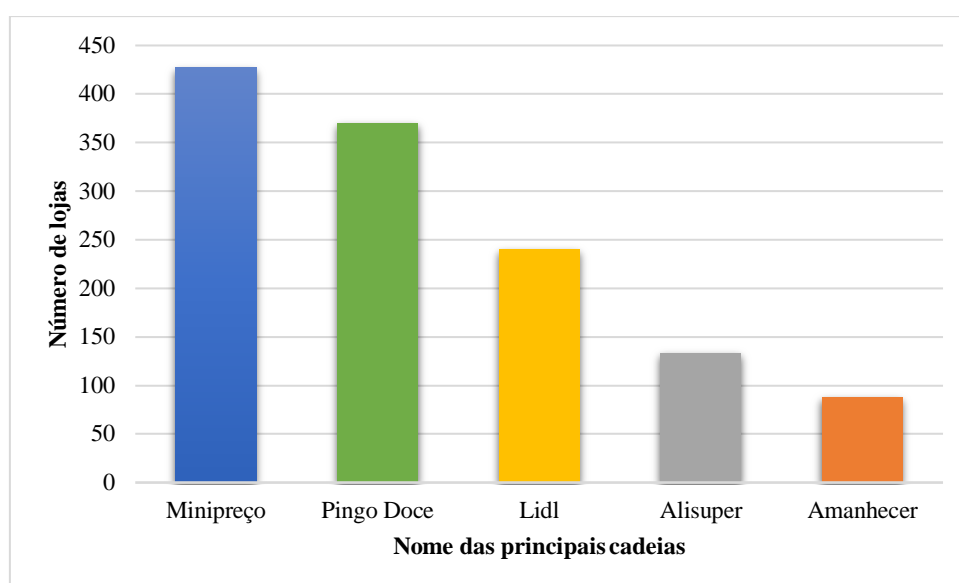


Figura 4.3: Concorrência às lojas Meu Super, com área até 540 m², no ano de 2015

Apesar de se apresentarem as cinco cadeias, apenas quatro delas foram consideradas concorrência direta às lojas em estudo, uma vez que as lojas Alisuper, não estão uniformemente distribuídas pelo território, havendo grande concentração desta cadeia na região do Algarve e poucas no resto do país. Assim apenas se consideraram como concorrência as lojas: Minipreço, Pingo Doce, Lidl e Amanhecer. Destas, apenas o LIDL é uma cadeia de lojas de desconto, tendo-se optado por a incluir neste leque de cadeias uma vez que estas lojas são capazes de atrair muitos clientes devidos aos preços praticados e não apenas pela proximidade aos clientes.

Depois de recolhidas as informações referentes a cada loja, analisou-se a sua distribuição na área de estudo. No Quadro 4.2, é possível observar quantas lojas de cada marca existiam em Portugal Continental. Para este projeto recolheram-se 1 511 lojas, das

quais 1 248 estavam em funcionamento, e as restantes foram encerradas pelas respetivas marcas. Da insígnia em estudo apenas 123 lojas estavam em funcionamento em 2015.

A marca com maior representatividade era o Minipreço, com cerca de 37% das lojas em estudo, em seguida do Pingo Doce com 33% e o LIDL com 17%. Com menor representatividade estavam os Meu Super, com 8% e os Amanhecer com 6% do total de lojas em estudo.

Quadro 4.2: Número de Lojas de cada insígnia em Portugal Continental

	Meu Super	Pingo Doce	LIDL	Minipreço	Amanhecer	Total
Em funcionamento	123	370	240	427	88	1248
Encerradas	9	43	3	115		170
Previstas		77	6	10		93
Total de Lojas	132	490	249	552	88	1511

Na Figura 4.4 pode observar-se a distribuição espacial das lojas Meu Super em Portugal Continental. Aqui consegue-se observar que existe uma maior concentração de lojas ao longo do litoral, com uma grande concentração de lojas no distrito de Lisboa, aproximadamente 29% das lojas desta insígnia (36 lojas, no total). As regiões com menor número de lojas eram o Algarve e Trás-os-Montes e Alto Douro. Regiões onde existiam apenas lojas, localizadas, respetivamente, em Lagos, na Praia da Rocha (Portimão), em Bragança e em Chaves.

Na Figura 4.5 é possível observar a distribuição espacial da concorrência das lojas Meu Super. Também nesta se nota uma maior concentração do número de lojas no litoral e mesmo sem a informação do quadro acima consegue-se observar que é a cadeia Minipreço a que tem maior implantação na área em estudo e o Amanhecer a com menor representatividade.

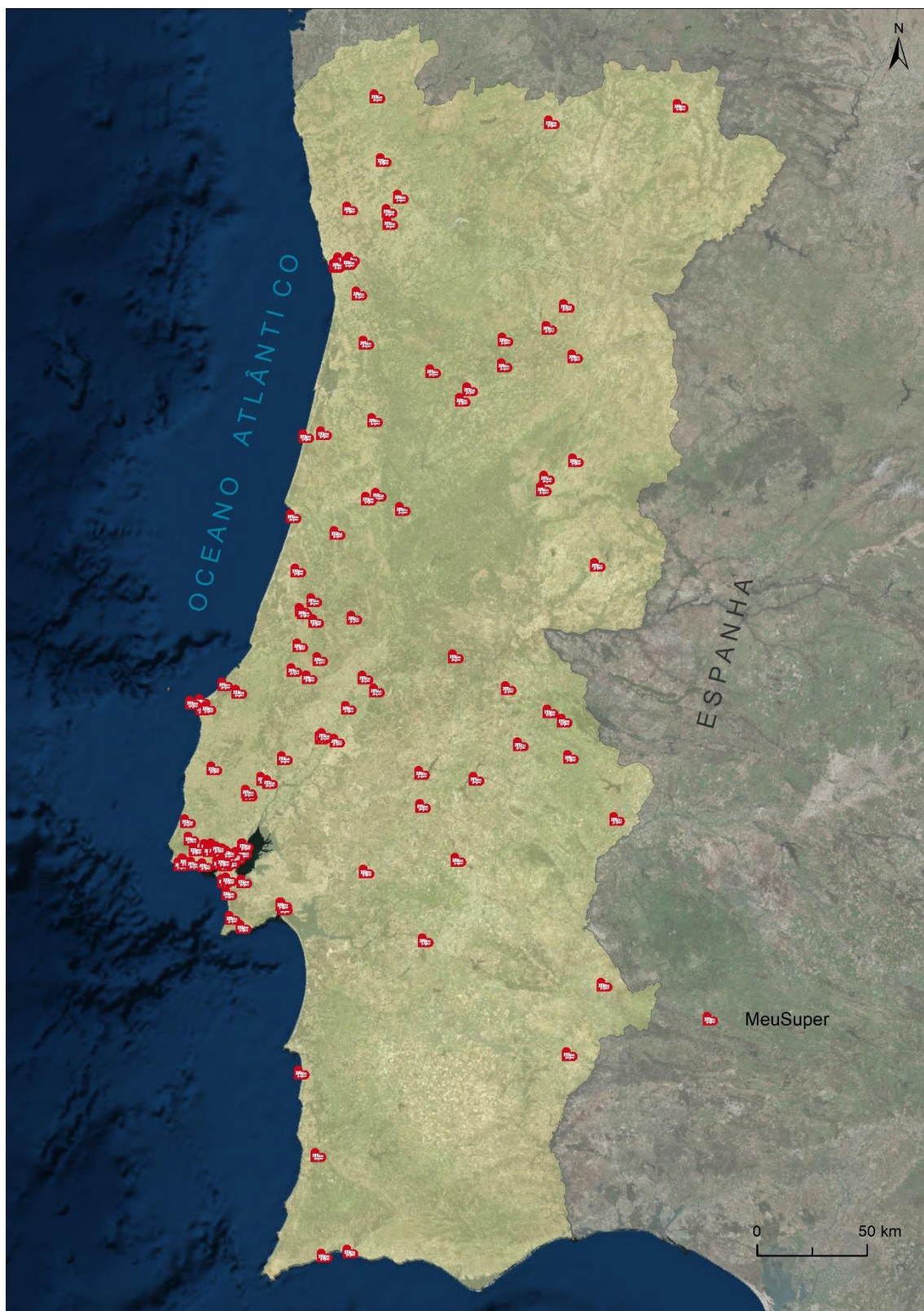
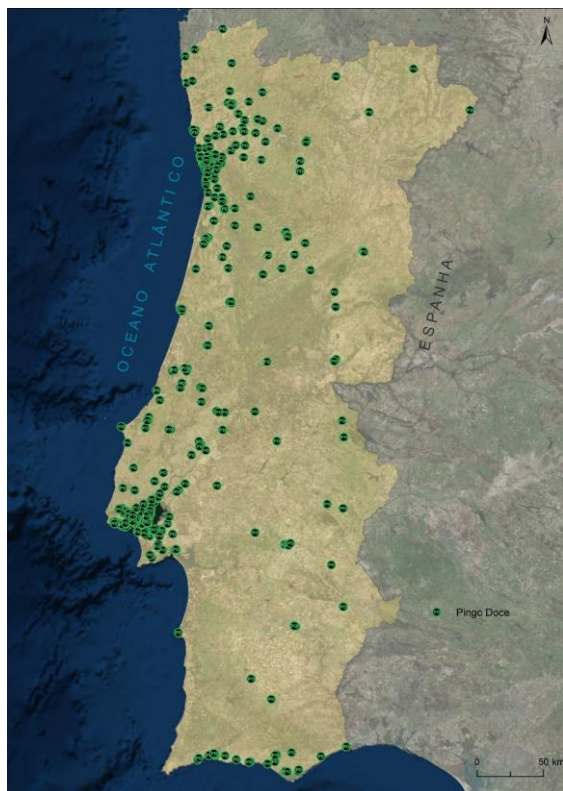


Figura 4.4: Distribuição das lojas da Insígnia Meu Super, no ano de 2015

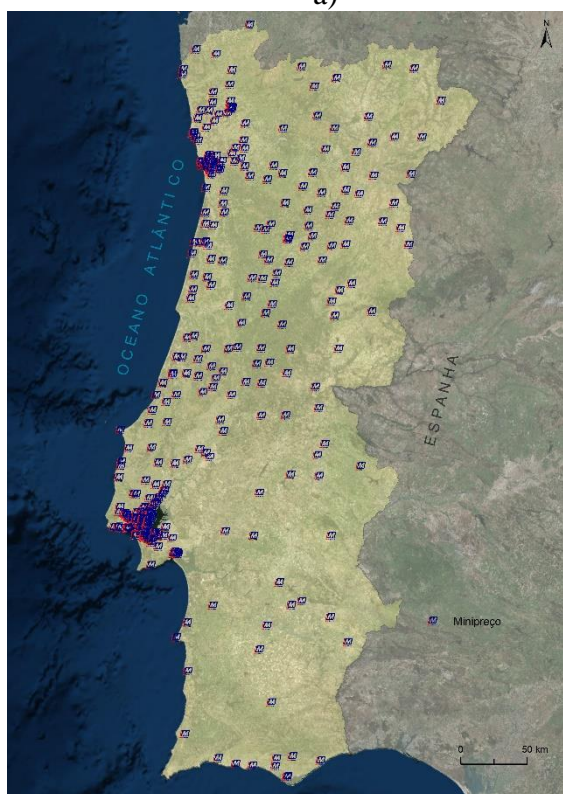
Base cartográfica: CAOP (IGP), World Imagery (ESRI)



a)



b)



c)



d)

Figura 4.5: Distribuição Espacial da Concorrência. a) Pingo Doce; b) Lidl; c) Minipreço; d) Amanhecer

Base cartográfica: CAOP 2016 (IGP), World Imagery (ESRI)

4.1.1. Difusão espacial

De modo a observar se existe alguma tendência de distribuição das lojas (concentração/dispersão, etc.), quer da insígnia em estudo quer as da concorrência, recorreu-se a uma análise do vizinho mais próximo (*Average Nearest Neighbour*). Esta ferramenta permite avaliar se a amostra se encontra agrupada (*clusters*), dispersa ou se a sua distribuição é aleatória no espaço.

O índice do vizinho mais próximo (IVMP) é um método de análise da distribuição espacial que determina o padrão de distribuição dos pontos de uma amostra calculando a distância média entre cada ponto e o seu vizinho mais próximo, independentemente da direção. Este foi criado e desenvolvido por dois botânicos, Philip Clark e Francis Evans, em 1954, para estudar as relações e medir padrões entre organismos-vivos, e, desde então tem sido utilizado nas mais diversas áreas, para os mais variados problemas (Cressie, 1991 citado por Levine, 2002).

O cálculo utilizado por esta ferramenta consiste no cálculo do valor médio da distância entre os centroides de todos os objetos (os próprios pontos) e os dos seus vizinhos mais próximos. Ebdon (1982) citado por Afonso (2018), afirma que antes de calcular o índice do vizinho mais próximo tem de se efetuar um cálculo intermédio onde se calculam as distâncias médias entre os pontos da amostra, sendo expresso da seguinte forma:

$$\bar{d}_{ran} = \frac{1}{2\sqrt{p}} \quad (4.1)$$

Onde \bar{d}_{ran} é o valor médio da distância entre os vizinhos mais próximos para um conjunto de pontos, e p é a densidade de pontos por unidade de área.

Feito este cálculo intermédio já se consegue calcular o índice de vizinho mais próximo onde a média da distância média ao vizinho mais próximo é calculada como a razão entre a distância média observada e a distância média esperada (Ebdon, 1982, citado por Afonso, 2018), sendo expressa da seguinte forma:

$$R = \frac{\bar{d}_{obs}}{\bar{d}_{ran}} \quad (4.2)$$

Em que R é o índice de Vizinho mais próximo (*Nearest Neighbor*), \bar{d}_{obs} é o valor médio da distância observada entre os vizinhos mais próximos e \bar{d}_{ran} é o valor médio da distância espectável entre os vizinhos mais próximos. O resultado deste índice revelará como se distribuem os pontos de amostra na área de estudo (Figura 4.6), dependendo do valor do índice de vizinho mais próximo (ESRI, sem data-b):

- Se o valor de IVMP for inferior a 1, a amostra é um cluster, ou seja, a amostra encontra-se aglomerada;
- Se o valor de IVMP for superior a 1, a amostra tende para a dispersão ou competição.

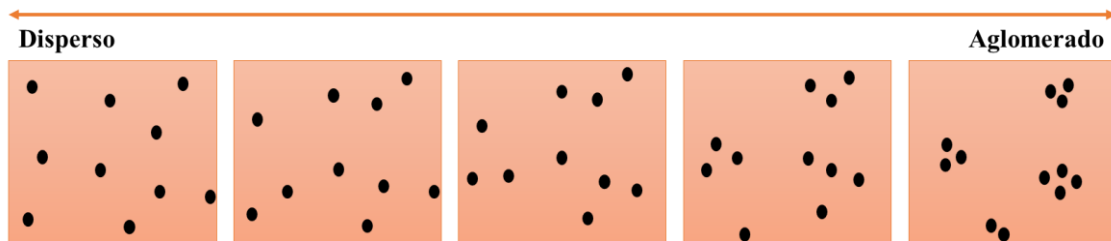


Figura 4.6: Escalas de distribuição de amostras

Fonte: ESRI (n.d.-b)

Dado que quando se analisou a distribuição das lojas em estudo, os IVMP foram todos inferiores a 1, pode-se afirmar que todas as cadeias apresentam um padrão de aglomeração (Quadro 4.3). Esta ferramenta ainda forneceu informações referentes ao *z-score* e *p-value* que são medidas de significância estatística que indicam se deve ou não rejeitar a hipótese nula, ou seja, rejeitar que os objetos sejam distribuídos aleatoriamente (ESRI, sem data-b). Estes parâmetros estão fortemente relacionados com o tamanho da área de estudo, uma vez que uma alteração no valor da área pode resultar em mudanças consideráveis nestes parâmetros, fazendo com que este índice seja mais eficaz para comparar diferentes objetos numa área de estudo fixa, como é o caso. Os valores referentes ao *z-score* estabelecem que o padrão de distribuição das lojas, quer da insígnia

‘Meu Super’ (Figura 4.7 a)) quer das insígnias da concorrência (Figura 4.7 b)), é concentrado, havendo menos de 1% de probabilidade que estes padrões de aglomeração (*clusters*) sejam produto do acaso (aleatório).

Quadro 4.3: Características do padrão de distribuição da loja Meu Super e da concorrência

	Meu Super	Pingo Doce	Lidl	Minipreço	Amanhecer
Índice do Vizinho mais próximo	0,771607	0,566399	0,853056	0,810894	0,706539
Z-score	-4,845810	-15,955919	-4,354992	-7,475686	-5,266511
P-value	0,000001	0,000000	0,000013	0,000000	0,000000
Distância Média Expectável (m)	13447,5265	7753,4341	9626,9639	7217,4042	15898,4076
Distância Média Observada (m)	10376,2046	4391,5389	8212,3425	5852,5469	11232,8409

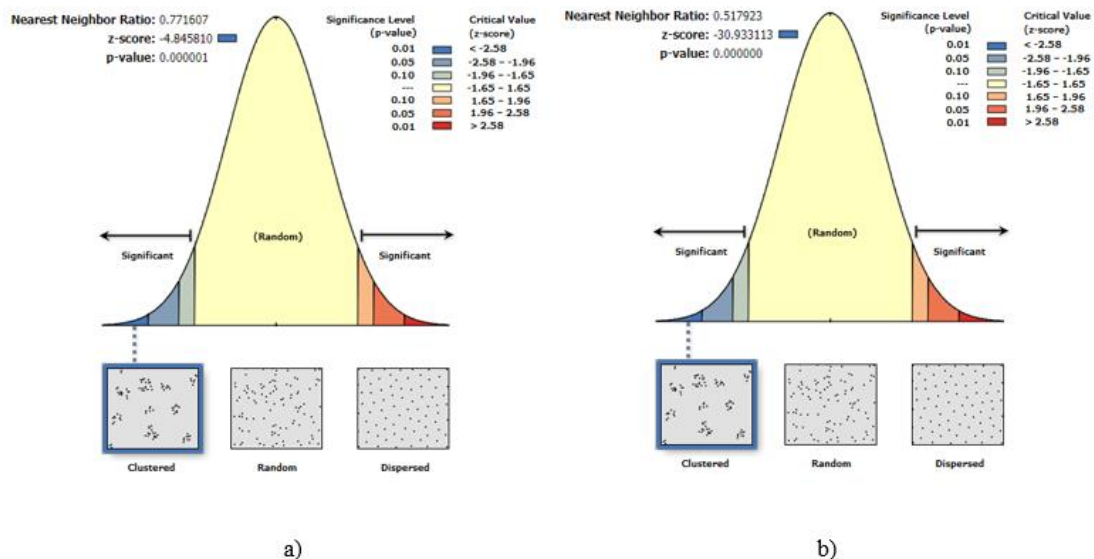


Figura 4.7: Distribuição espacial da loja Meu Super e da concorrência

4.1.2. Caracterização das lojas Meu Super

Particularizando a insígnia em estudo, o Meu Super, esta consiste num *franchising* de estabelecimentos alimentares de do grupo Sonae MC (Sonae, sem data), implantado em Portugal Continental e Arquipélagos. As lojas têm uma dimensão entre os 52 m², em Campo de Ourique e Marvila, e os 540 m² em Caxarias. A primeira loja a ser aberta, e ainda em funcionamento, foi o MS do Estoril (23/07/2011) e a mais recente é a de Albarraque (30/12/2014). Com a criação das lojas Meu Super, o grupo Sonae passou a ter maior presença em áreas residenciais ou de serviços com elevado tráfego pedonal. Deste modo, o grupo Sonae conseguiu disponibilizar aos seus clientes produtos de Marca Continente, aproveitando a proximidade proporcionada por estes estabelecimentos, e os franchisados obtiveram apoio na gestão das lojas, um preço competitivo, o acesso a marcas próprias do Continente e a produtos de outros fornecedores (Sonae, 2015).

Como referido anteriormente, foi possível obter dados referentes ao número de transações e vendas brutas efetuadas por esta insígnia, por loja, com e sem cartão de fidelização, a dimensão das lojas, a sua data de abertura, se uma loja já era uma mercearia antes de ser um MS ou se foi criada de raiz e se há facilidade de estacionamento. Todas estas variáveis foram recolhidas para os meses de março, abril e maio (apenas até dia 13 de 2015). Na Figura 4.8, pode observar-se o total de vendas brutas e transações com e sem cartão de fidelização, entre abril e maio de 2015.

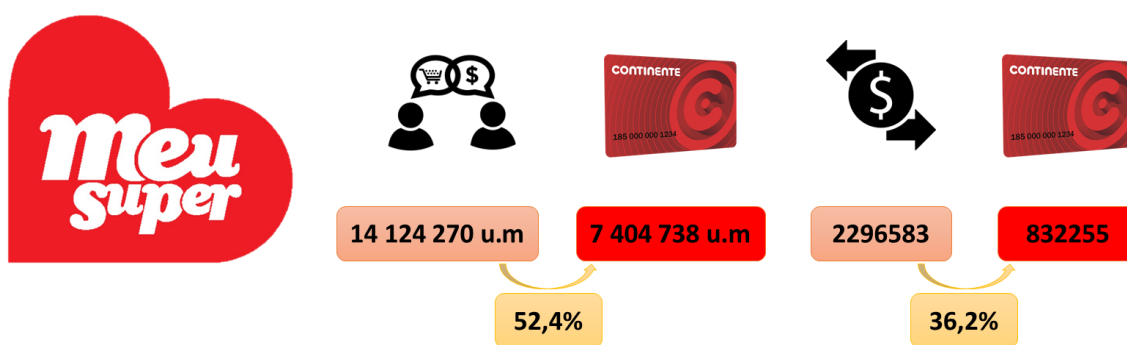


Figura 4.8: Volume de vendas e transações das lojas Meu Super, de abril a maio de 2015

Durante estes meses o MS fez um total de 2 296 583 transações comerciais, dos quais apenas 36,2% dizem respeito a transações com cartão de fidelização (832 255). Essas transações corresponderam a um total de 14 124 270 unidades monetárias de vendas brutas, em que 52,4% desse valor dizem respeito a vendas com a utilização do

cartão de fidelização (7 04 738 u.m.). Entre os meses de março e abril assistiu-se a um crescimento, quer da percentagem de vendas brutas fidelizadas quer das transações fidelizadas, havendo, no entanto, um decréscimo no mês de maio de 1% e 7%, respetivamente. Este decréscimo no número de vendas brutas e transações totais (em anexo), no mês de maio, deve-se ao facto da informação se reportar apenas aos primeiros 13 dias do mês.

Das transações efetuadas durante os três meses, as lojas de Arcos de Valdevez, Moreira de Cónegos, Amareleja, Calvão e Arrabal apresentaram menor número de transações comerciais (Figura 4.9). No mesmo período, as lojas que apresentaram maior número de transações comerciais foram as de Almeirim, Parque das Nações, Camarinha, Cascais e Moscavide (Figura 4.10).

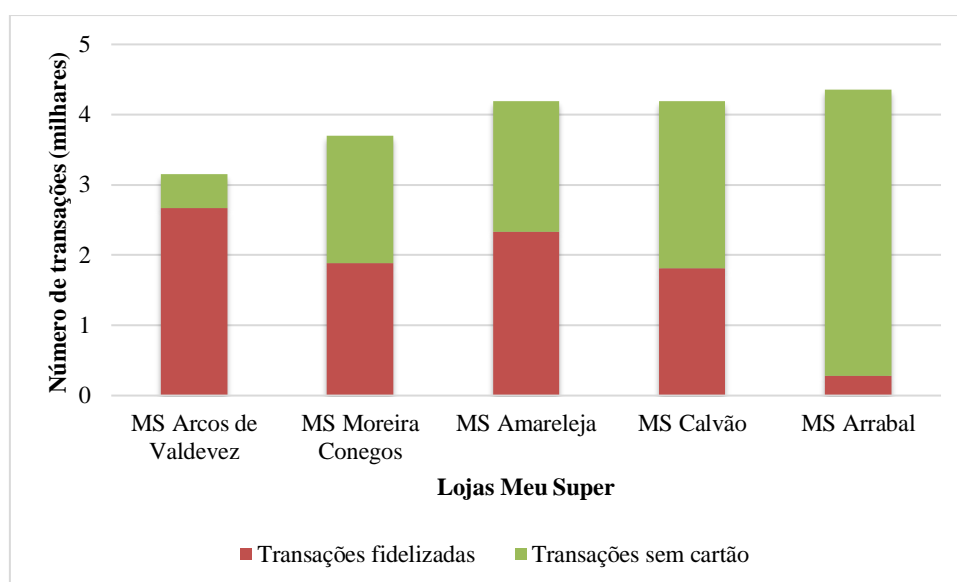


Figura 4.9: Lojas MS com menos transações comerciais trimestrais

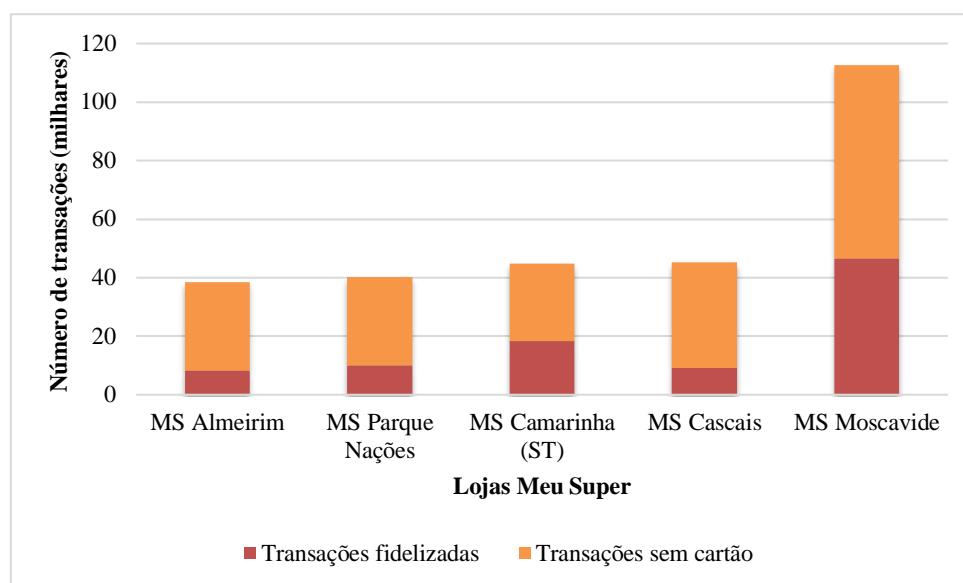


Figura 4.10: Lojas MS com mais transações comerciais trimestrais

No conjunto das 5 lojas com menor volume de transações comerciais, as três lojas com menor número de transações (Figura 4.9) são as que apresentam mais transações com cartão de fidelização, esta situação pode dever-se ao facto de serem áreas com população idosa e com baixo poder de compra que recorre à loja mais próxima de si e utiliza o cartão de fidelização para aproveitar as vantagens fornecidas por este (descontos e promoções). Relativamente às lojas com mais transações comerciais (Figura 4.10), grande parte das transações efetuadas não são feitas com cartão de fidelização. Existe uma diferença de aproximadamente 11 000 transações entre as lojas de Arcos de Valdevez e de Moscavide, sendo que, para o período dos dados disponibilizados, a loja de Moscavide contribuiu com, aproximadamente, 5% do total de transações comerciais da insígnia Meu Super e a loja de Arcos de Valdevez apenas 0,1% do total de transações.

Com os dados disponibilizados, foi também possível observar que as lojas que apresentam menos vendas brutas, nos três meses, foram os Meu Super de Moreira de Cónegos, Cabeço de Montachique, Calvão, Abrunheira e Oeiras (Figura 4.11). Os que apresentavam mais vendas brutas foram o Meu Super de Moscavide, Cascais, Parque das Nações, Montargil e Arroios (Figura 4.12).

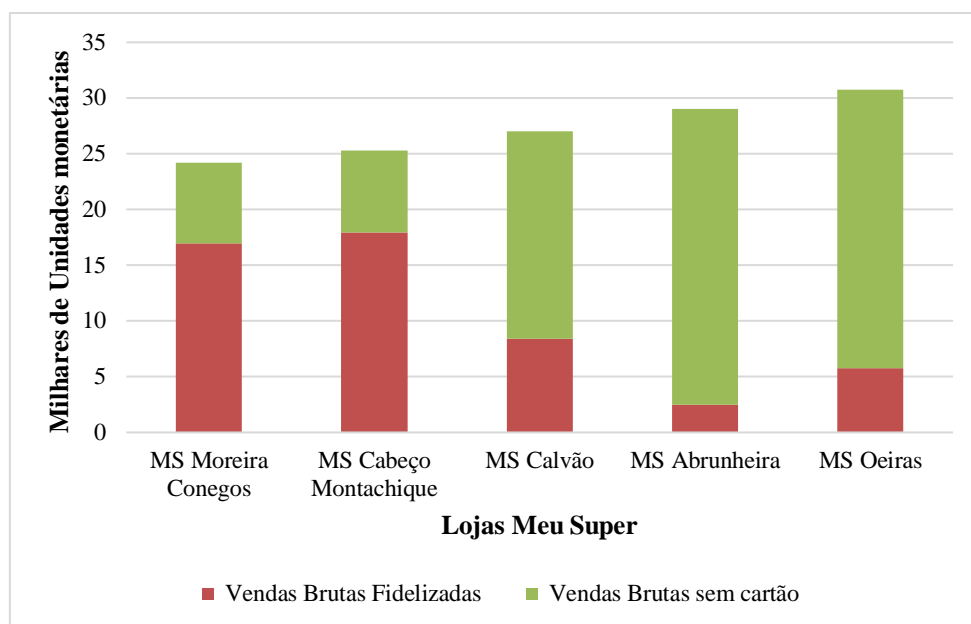


Figura 4.11: Lojas MS com menos vendas brutas trimestrais

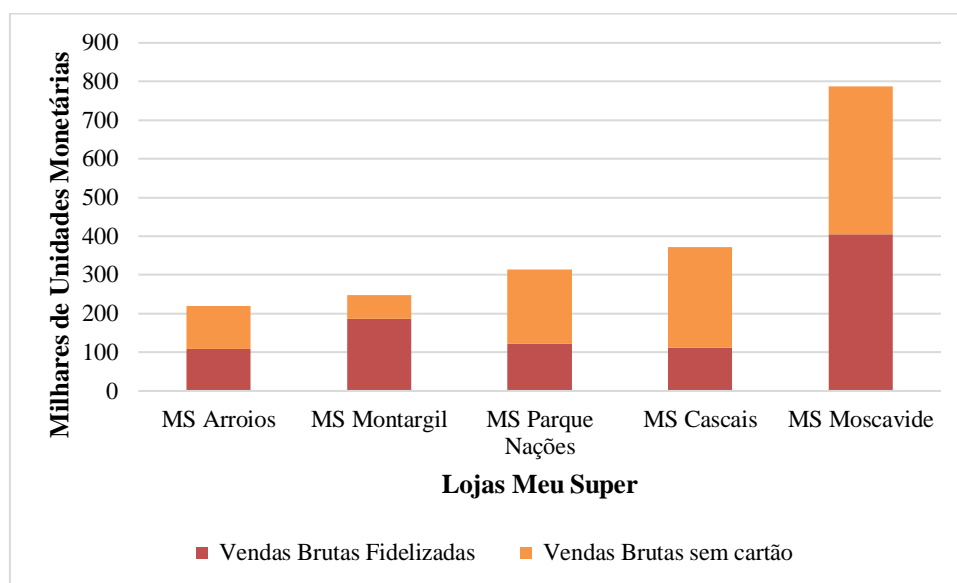


Figura 4.12: Lojas MS com mais vendas brutas trimestrais

Quando se observa a Figura 4.11, é possível notar que são nos dois MS que apresentam menos vendas brutas totais onde existem mais vendas fidelizadas, cerca de 10 000 u.m. a mais do que as restantes lojas com menos vendas. Estas lojas localizam-se em áreas em que o poder de compra não é muito elevado e que, por isso, a população recorre às promoções que o cartão de fidelização fornece, acabando por explicar o valor mais elevado de vendas brutas fidelizadas. No caso das lojas com mais vendas (Figura 4.12), são os MS de Montargil e Moscavide que apresentam mais vendas fidelizadas que,

tal como as duas lojas com menor vendas, estão localizadas em bairros com um poder compra reduzido em que a população recorre aos descontos e promoções do cartão de fidelização. Relativamente às restantes três lojas, estas situam-se em áreas em que o poder de compra é superior às restantes e, por isso, poderá não haver tanta necessidade de se utilizar as vantagens proporcionadas pelo cartão de fidelização, o que explica o facto de que, nestas lojas, a maior parte das vendas sejam sem cartão de fidelização.

Quando se analisa pormenorizadamente a evolução do crescimento de vendas brutas fidelizadas para as duas lojas com mais e menos vendas, comprova-se o que foi referido anteriormente, que apesar de terem os valores totais de vendas brutas mais baixos são as duas lojas com menos vendas que apresentam uma maior taxa de vendas fidelizadas, ou seja, do valor total de vendas a maior parte delas provêm de clientes com cartão de fidelização.

Através da Figura 4.13, que representa os gastos médios dos clientes com o uso do cartão de fidelização, pode notar-se que nas lojas quer com mais vendas brutas (representados com estrelas), quer com menos (representados com pontos), a média de gastos dos clientes não ultrapassa as 200 unidades monetárias. Também é possível observar que os municípios com a média de gastos mais elevados nas lojas Meu Super são Vila Nova de Paiva e Sines. Ou seja, nestas lojas é provável que a população efetue regularmente as suas compras mensais nestas lojas, uma vez que ultrapassam o valor médio de gastos no país.

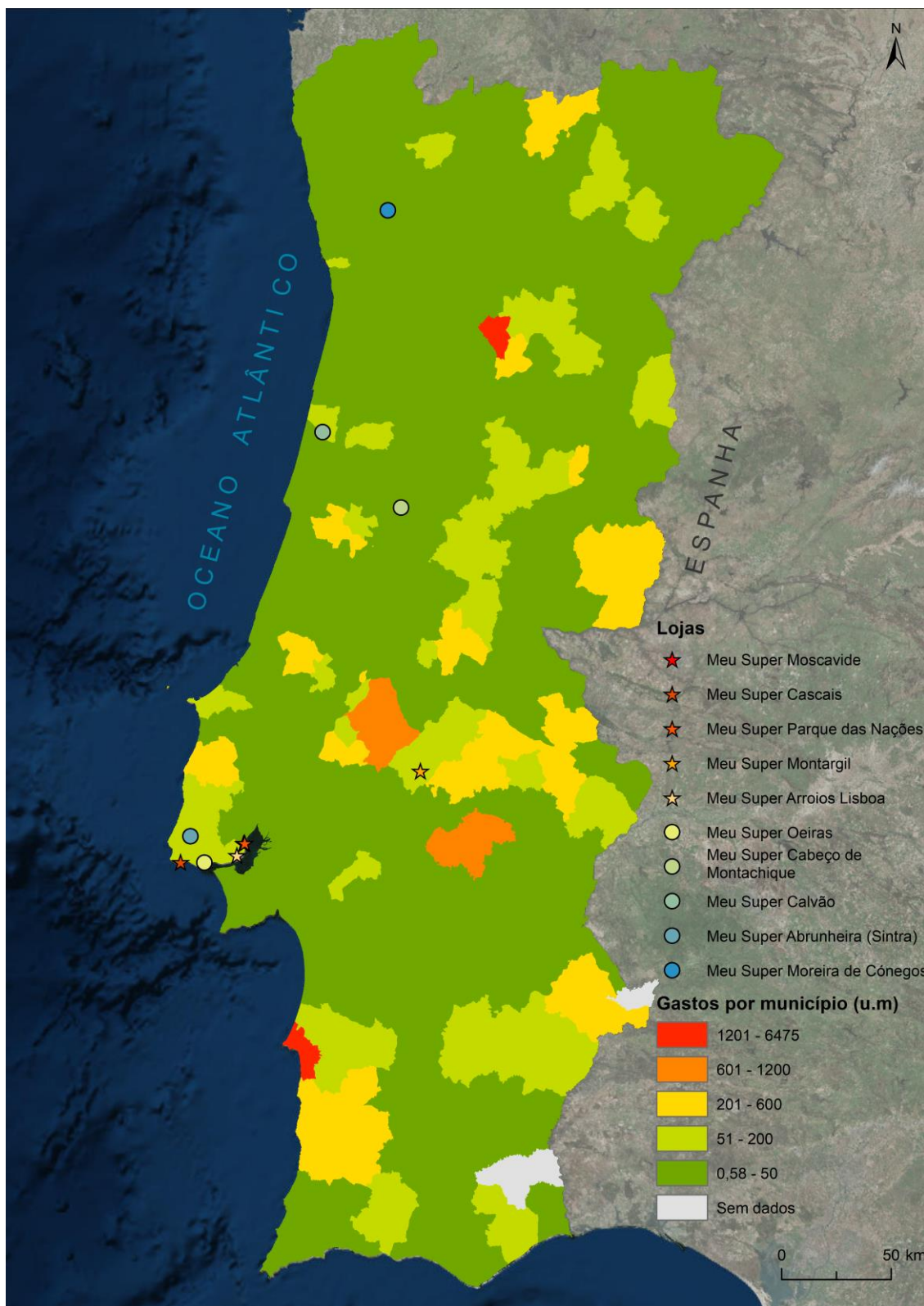


Figura 4.13: Média de Gastos, com cartão, por município
 Base cartográfica: CAOP (IGP), World Imagery (ESRI)

4.2. Caracterização da Procura

Os estudos sobre o comportamento dos clientes de lojas de venda a retalho lidam sobretudo com a identificação dos clientes e dos seus padrões de comportamento de consumo, sendo que os objetivos dos estudos tendem a responder a questões como ‘quem compra’, ‘o que compra’, ‘onde compra’, ‘quando compra’ e ‘como compra’ (Applebaum, 1951, p.172).

Hoje em dia é cada vez mais complexo definir um perfil de cliente, uma vez que estes se encontram em áreas cada vez mais diversificadas (Catunda, 2010) e com hábitos de consumo variados. De acordo com Copeland (1923), o perfil de cliente de lojas de retalho varia consoante as categorias de bens que existem, nomeadamente os *convenience goods*, *shopping goods* e os *specialty goods*. Os *convenience goods* são os bens que são comprados de forma regular, sem um tipo de procura específica ou que necessite de comparação com outros bens. Os *shopping goods* são os bens em que os consumidores fazem uma pesquisa prévia antes de os adquirirem, de modo a obter o melhor preço. E, por último, os *specialty goods* são aqueles em que a escolha dos consumidores é determinada pelo nome da marca. Sendo que o objetivo é abrir mais estabelecimentos de comércio alimentar de proximidade, pode afirmar-se que estes correspondem à primeira categoria de bens, uma vez que a escolha por estes estabelecimentos deixou de ser tão pensada e é feita, cada vez mais, de acordo com a oportunidade (Desse, 2001 *citado por* Afonso, 2018, p.79).

Nesta fase, houve a necessidade de definir um perfil de cliente. No entanto, são várias as características que influenciam a definição de consumidores e que têm impacto na decisão por uma loja em detrimento de outras. Entre algumas delas estão: o género, a idade, a etnia, a religião, o grau de instrução, a ocupação profissional ou a classe económica dos clientes (Applebaum, 1951, p.172-173). No entanto, para obter o perfil do cliente exacto torna-se necessário conhecer as principais características dos clientes deste tipo de lojas e como é que estes se comportam. Assim sendo foi necessário recorrer a inquéritos.

De forma a conhecer o tipo de clientes e as dinâmicas das lojas em estudo obteve-se acesso a dados referentes a inquéritos realizados pela Sonae, sobre as lojas de proximidade Meu Super em 2015. Entre as informações recolhidas encontram-se as disponibilizadas na Figura 4.14, e, ainda, informações referentes às lojas que costumam

visitar, motivos que os levaram a visitar a loja, de onde e para onde se dirigiam quando visitaram a loja e o grau de satisfação perante a loja.

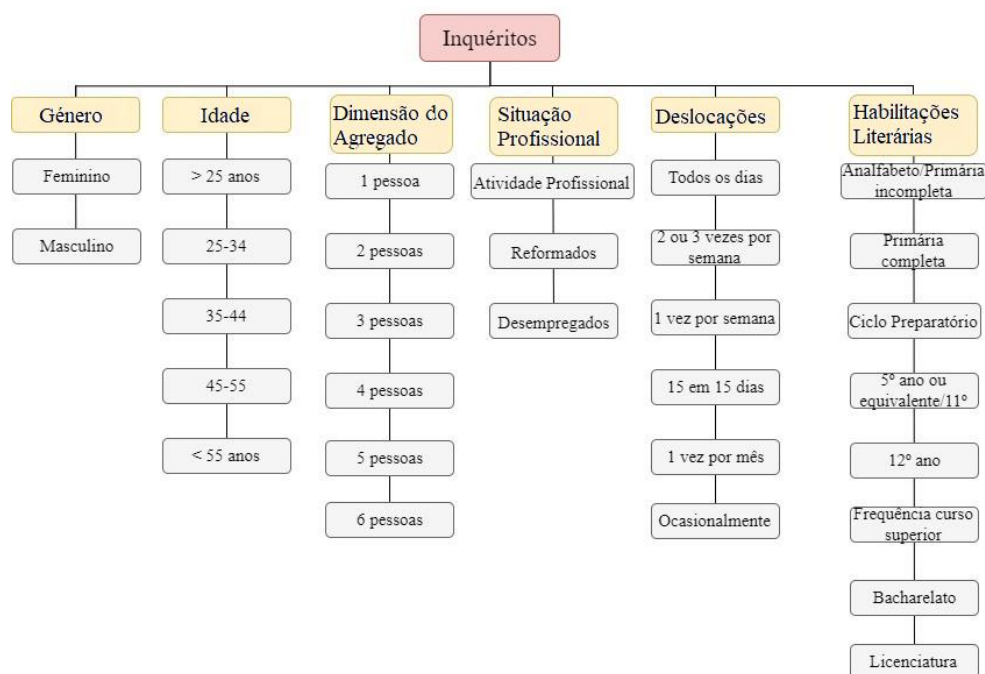


Figura 4.14: Informações retiradas dos inquéritos realizados pela Sonae em 2015

Foram realizados cerca de 100 inquéritos em oito lojas distintas, perfazendo uma amostra total de 861 clientes inquiridos. Desta forma, foi possível segmentar a população e ter uma ideia do perfil de cliente das lojas em estudo. As principais conclusões são mencionadas na Figura 4.15.

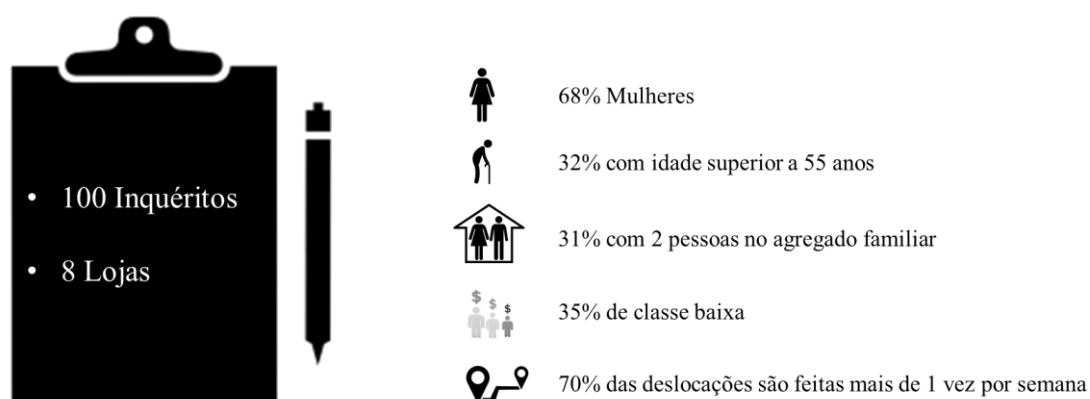


Figura 4.15: Resultados dos inquéritos

A partir dos inquéritos realizados, Afonso (2018) afirma que o perfil de cliente em nada era influenciado pela localização da loja e que as tipologias eram relativamente semelhantes no que toca à frequência de compra, forma de deslocação, ponto de partida e de chegada. As principais conclusões que se retiraram dos inquéritos foram que do total de clientes inquiridos 68% eram do sexo feminino (463 clientes) e os restantes 32% eram do sexo masculino (218); dos grupos etários definidos, 32% dos clientes apresentava idade superior a 55 anos; cerca de 35% dos inquiridos eram de classe baixa (rendimento até 800€, nível de escolaridade baixo, trabalhadores não qualificados e/ou pessoal de serviço ou vendedores) e, aproximadamente 31% dos clientes pertencia a um agregado familiar constituído por duas pessoas. Ainda foi possível descobrir que cerca de 70% das deslocações são feitas mais do que uma vez por semana e são feitas, na sua maioria, a pé, o que vem reforçar a ideia de lojas de bairro, em que as deslocações se fazem de casa ou a caminho de casa, não sendo necessário grandes deslocações para adquirir os produtos.

Com estas informações, foi possível, através dos censos, caracterizar o território em termos demográficos. Os Censos de 2011, realizados pelo INE, são uma grande base de dados demográficos que, juntamente com os dados sobre a localização dos supermercados em funcionamento, permitiram a realização deste projeto. No entanto, nem todas as informações inseridas na Base Geográfica de Referenciação de Informação foram necessárias, havendo necessidade de tratar a informação para melhor adequação ao projeto, ou seja, foram tratados de acordo com os inquéritos realizados pela empresa.

No Quadro 4.4, estão representadas as variáveis que foram utilizadas neste projeto. Estes dados têm informação referente ao género, idade, constituição das famílias, e, como não existiam dados socioeconómicos, tentou encontrar-se uma solução que pudesse refletir as diferenças no poder de compra da população residente e que caracterizassem a classe económica da população.

Assim, com o propósito de construir o target de mercado, nomeadamente a população residente que faça compras regularmente em locais próximos de casa onde se podem deslocar a pé, recorreu-se à BGRI, particularmente às variáveis referidas no Quadro 4.4, para caracterizar a população residente em Portugal Continental.

Quadro 4.4: Dados provenientes dos censos 2011

Perfil Médio dos inquiridos		Informações correspondentes aos inquéritos e complementares nos censos 2011	
Variáveis Demográficas	Mulheres	População Residente (Mulheres)	
		População Residente (Homens)	
	Residentes com idade superior a 55 anos	<u>Residentes com idades entre:</u>	
		- 0 e os 14 anos	
		- 14 e os 19 anos	
		- 20 e os 64 anos	
		- Mais de 64 anos	
	Famílias com 2 pessoas no agregado familiar	<u>Famílias clássicas com:</u>	
		- 1 ou 2 elementos	
		- 3 ou 4 elementos	
- 5 ou mais elementos			
Variáveis Económicas	Classe baixa	Indivíduos residentes empregados	
		Indivíduos residentes reformados ou pensionistas	
		Indivíduos residentes sem atividade profissional	
		Indivíduos residentes desempregados à procura de novo emprego	
		1º Ciclo completo	
		2º Ciclo completo	
		3º Ciclo completo	
		Ensino Secundário completo	
		Ensino Superior	
		Sector Primário	
		Sector Secundário	
		Sector terciário	
Target: População Residente			

No entanto, na modelação geográfica é habitualmente difícil obter dados adequados e concretos ao objetivo e, sendo que para atingir o objetivo é fundamental ter o maior rigor possível, decidiu-se recorrer à cartografia dasimétrica, para delimitar com maior exatidão a distribuição da população.

4.2.1. Cartografia Dasimétrica

Como referido anteriormente (capítulo 4), os dados detalhados sobre a população em Portugal provêm dos censos nacionais, em que a menor unidade territorial cartografada corresponde a quarteirões/subsecções. No entanto, estas áreas não são homogéneas no espaço, fazendo com que também não sejam homogéneas no número de habitantes entre as áreas urbanas e áreas urbanas-rurais (Garcia et al., 2016). Por norma, as áreas urbanas-rurais apresentam maior área do que as urbanas, no entanto apresentam menor número de população. De modo a obter o maior rigor possível no cálculo da procura/clientes por estabelecimentos comerciais de proximidade, decidiu-se recorrer à cartografia dasimétrica.

A cartografia dasimétrica é uma técnica cartográfica para representar dados de valores quantitativos utilizando limites que delimitam a área mapeada em zonas homogéneas com o objetivo de melhor retratar a superfície estatística subjacente (Eicher & Brewer, 2001). Isto é, é uma técnica evolutiva do mapa coropleto, que permite utilizar informação geográfica auxiliar para desagregar e melhorar a representação do fenómeno (Silva, 2009), no caso desta dissertação, a distribuição da população, limitando-o a áreas em que o fenómeno pode ocorrer, nomeadamente, as áreas urbanas.

A cartografia dasimétrica foi concebida como um tipo de mapa temático, tal como os mapas coropletos, no século XIV, para o estudo da distribuição populacional, no entanto neste século ainda não era designada com este termo. A partir do século XX, quando o geógrafo russo Benjamin Semenov-Tian-Shansky utilizou o termo “dasimétrico” pela primeira vez (1911), começou-se a diferenciar os mapas coropletos dos dasimétricos, tornando-se os primeiros os mais populares na cartografia moderna (Mennis, 2009). As principais diferenças entre estes dois tipos de cartografia prendem-se com o facto de os limites zonais nos mapas dasimétricos serem baseados em mudanças acentuadas na superfície estatística, enquanto os limites zonais nos mapas coropletos delimitam unidades de terreno estabelecidas para fins mais gerais; as zonas dasimétricas

são desenvolvidas para serem internamente homogéneas, enquanto as coropletas apresentam vários níveis de homogeneidade interna, uma vez que não são demarcadas com base nos dados; e os métodos de mapeamento da cartografia coropleta tornaram-se padronizados enquanto existe uma ampla gama de procedimentos dasimétricos (Eicher & Brewer, 2001).

O aparecimento dos sistemas de informação geográfica e da deteção remota conseguiu trazer inovações e renovar interesses nesta técnica cartográfica (Eicher & Brewer, 2001). Os SIG concederam as ferramentas necessárias para a sobreposição de variáveis (atributo necessário para a cartografia dasimétrica), e a deteção remota forneceu um grande recurso de dados auxiliares necessários para realizar mapeamentos dasimétricos da população, particularmente sob a forma de mapas de cobertura de terra classificados (Mennis, 2009).

Para este projeto utilizou-se o mapa de ocupação do solo para se delimitar as áreas urbanas, uma vez que será nestes locais onde existirá maior concentração de população e, por isso, onde será mais relevante colocar um estabelecimento comercial de proximidade. Para isso, selecionaram-se as áreas correspondentes ao ‘Território Artificializado’ (exceto as áreas de extração de inertes, áreas portuárias e aeródromos) e exportou-se como ‘Áreas Urbanas’. Seria interessante ter áreas mais pormenorizadas da localização da população, como por exemplo o edificado, no entanto para esta escala de trabalho não foi possível obter a variável com o rigor necessário. Através da Base Geográfica de Referência de Informação (BGRI), obteve-se acesso aos censos da população portuguesa no ano de 2011. Com esses dados calculou-se a área em km² de cada subsecção estatística para, em seguida, se conseguir calcular a densidade populacional (em anexo). Em seguida, e de modo a obter as áreas urbanas em cada subsecção estatística, recortou-se a BGRI por estas áreas e tornou-se a calcular a área, desta vez das áreas urbanas dentro das subsecções, e a densidade populacional. Por fim, para obter a distribuição dasimétrica da população fez-se os residentes pelo rácio entre a área de cada área urbana e a área total da subsecção (em anexo). A este processo dá-se o nome de Cartografia Dasimétrica (Figura 4.16).

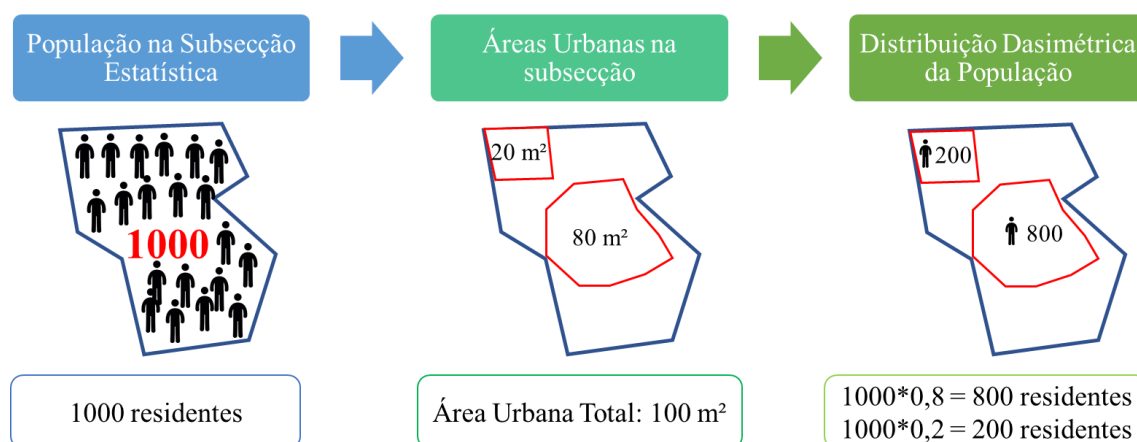


Figura 4.16: Modelo de cartografia dasimétrica.

Na Figura 4.16 está um exemplo do processo da cartografia dasimétrica. As subsecções estatísticas são a menor unidade de terreno disponível para a representação da população em Portugal, no entanto esta distribuí uniformemente a população, como demonstra a figura. Com esta técnica cartográfica, pretendeu-se limitar a presença de população às áreas urbanas, deste modo a população apenas será distribuída pelas áreas urbanas. No exemplo da Figura 4.16, os 1000 residentes que estavam distribuídos uniformemente pela subsecção vão ser circunscritos a 100 m², correspondentes às duas áreas urbanas presentes. De modo a distribuir a população multiplicou-se a população residente pela importância de cada área urbana no total das áreas urbanas, obtendo-se o total de residentes por área urbana. A Figura 4.17 representa um exemplo real da distribuição dasimétrica da população para a área em estudo.

Após a concretização do modelo, foi possível observar as diferenças entre a distribuição da população nas subsecções e nas áreas urbanas. Na Figura 4.17 foram assinaladas e comparadas duas áreas no município de Sintra. Pode observar-se que quando se opta pelo cálculo com base nas áreas urbanas, a densidade populacional é mais elevada do que quando se opta pelo total da subsecção estatística. Essas diferenças acentuam-se em subsecções em que a área urbana não ocupa a totalidade da subsecção, como é o caso do local A. A discrepância no valor da densidade populacional deve-se ao facto de apenas 68.6% da subsecção A poder apresentar população residente (área correspondente à área urbana), em vez da totalidade da subsecção estatística. No caso do local B, como a área urbana corresponde a toda a subsecção, não existem alterações nos valores.

	Local A	Local B
População residente (hab.) - Subsecção	73	63
Área da subsecção (km2)	0,046	0,013
Densidade populacional na subsecção (hab./km2)	1553,24	4990,31
Área Urbana (km2)	0,032	0,013
Densidade populacional na área urbana (hab./km2)	2264,47	4990,31



Figura 4.17: Exemplo da distribuição dasimétrica da população

Base cartográfica: CAOP 2016 (IGP), BGRI (INE), World Imagery (ESRI)

Adaptado de Garcia, 2012

A utilização da cartografia dasimétrica em detrimento do uso, de apenas, das subsecções representadas na BGRI traz a vantagem de obter um número mais realístico da população residente, uma vez que os nos censos o número de habitantes é , por norma, sobrestimado (Garcia et al., 2016). Para este projeto decidiu-se utilizar os dados em estrutura vetorial, uma vez que deste modo serão conservadas as “fronteiras” entre os quarteirões permitindo uma melhor delimitação, uma vez que se se optasse por estrutura matricial, os limites/fronteiras seriam mais grosseiros, não representando com exatidão a área urbana.

Apesar de se ter referido a residentes, esta técnica também foi empregue para outras variáveis disponíveis nos censos da população, nomeadamente as referidas no Quadro 4.4. Estas variáveis referem-se às informações retiradas dos inquéritos realizados pela Sonae aos clientes das lojas do Meu Super, referidos anteriormente no capítulo Caracterização da Procura.

4.2.2. Geodemografia

De seguida serão apresentados os vários mapas com a distribuição espacial da população, referentes às variáveis mencionadas no Quadro 4.4. As variáveis mencionadas foram transformadas em estrutura matricial com uma resolução de 100x100, uma vez que

deste modo garante-se uma maior qualidade na visualização, apesar de se perder na exposição dos dados/delimitação das áreas. Em seguida, foram classificadas segundo os valores de um desvio-padrão, sendo possível, deste modo, observar quanto o valor de atributo de uma característica varia da média, criando quebras de classe com intervalos de valores iguais de um desvio-padrão.

A primeira variável refere-se ao *target* deste projeto, nomeadamente, os indivíduos residentes na área de estudo. Nas áreas urbanas em Portugal residem 9 937 282 indivíduos, que se distribuem como mostra a Figura 4.18. O número máximo de indivíduos residentes numa área urbana é de 1 742, sendo que estas áreas se localizam maioritariamente nas grandes metrópoles portuguesas (Lisboa e Porto). Através da Figura 4.18, observa-se que os valores mais elevados de população residente se encontram ao longo do litoral e, assiste-se a uma polarização das áreas metropolitanas de Lisboa e Porto. No interior do país é possível observar que as maiores aglomerações de população residente se encontram, na sua maioria, nas sedes de município e ao longo das vias de comunicação.

Visto que nos inquéritos realizados pela Sonae mais de metade dos visitantes das lojas eram mulheres será interessante observar como estas se distribuem, na Figura 4.19 observa-se a distribuição das mulheres e homens nas áreas urbanas portuguesas. É de salientar que existem mais 448 312 mulheres residentes do que homens em Portugal, chegando a haver áreas urbanas com 934 mulheres residentes (Figura 4.19 a)) enquanto que o máximo de homens por área urbana chega apenas aos 841 residentes (Figura 4.19 b)). A distribuição, quer das mulheres quer dos homens, apresenta semelhanças, uma vez que as maiores aglomerações se encontram nas áreas metropolitanas de Lisboa e Porto e em algumas cidades no litoral (como Almada ou Faro); e é no interior onde se encontram as áreas urbanas com menor número de indivíduos.

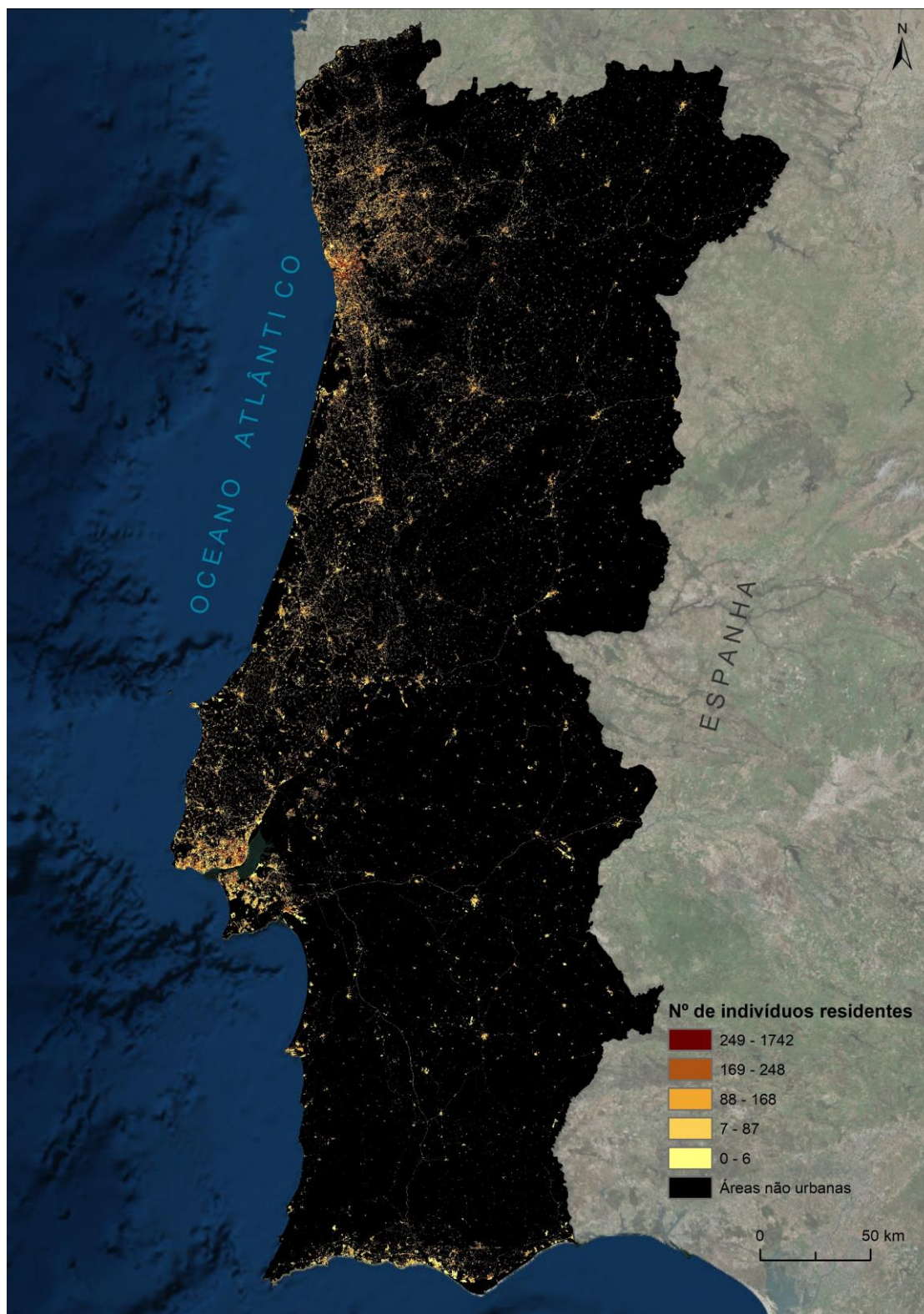


Figura 4.18: Distribuição do número de indivíduos residentes, em 2011

Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

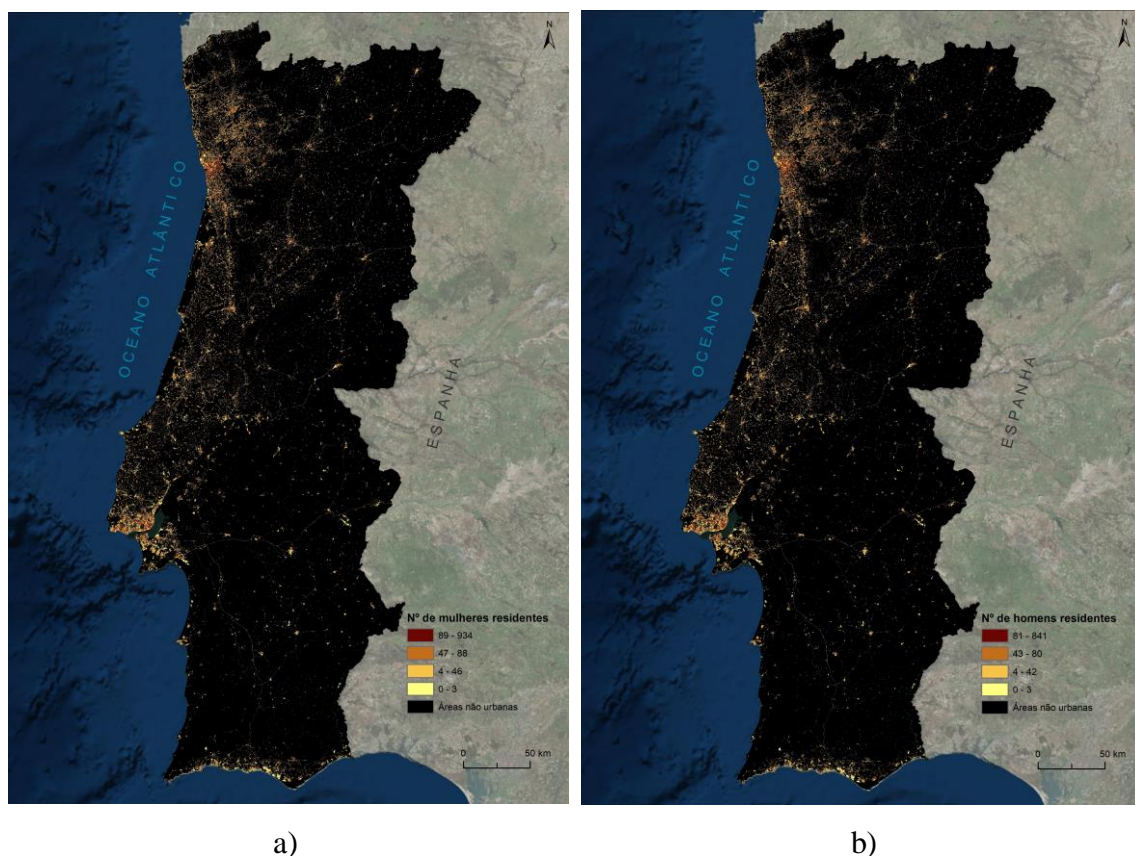


Figura 4.19: Distribuição do número de mulheres (a) e homens (b) residentes, em 2011

Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

A população residente está organizada em diferentes estruturas etárias, no entanto, apenas os indivíduos que podem fazer compras e que tenham poder de compra são relevantes para este estudo. Assim sendo, na Figura 4.20 observa-se a distribuição do número de indivíduos residentes entre os 20 e os 64 anos (a) e com mais de 64 anos (b). Nestas estruturas etárias estão a população adulta com poder de compra que pondera sobre o melhor local para efetuar as suas compras do mês e que, por vezes, recorrem às lojas de proximidade numa situação de aperto; e a população que apresenta maior probabilidade de recorrer às lojas de proximidade, de modo a não percorrer grandes distâncias para fazer as suas compras.

É no primeiro segmento onde se encontra a população em idade laboral e onde, dependendo do poder de compra, se opta por determinado estabelecimento para efetuar as suas compras. Estes, tendencialmente, efetuam as suas compras semanalmente a caminho de casa, em lojas de pequena/média dimensão, ou as compras do mês numa loja

de grande dimensão. No segundo segmento, a população com mais de 64 anos, como já apresenta outros hábitos de consumo, tendencialmente, efetua as suas compras diariamente, recorrem a lojas de proximidade por estas conseguirem trazer até si produtos frescos e com qualidade sem a necessidade de uma grande deslocação.

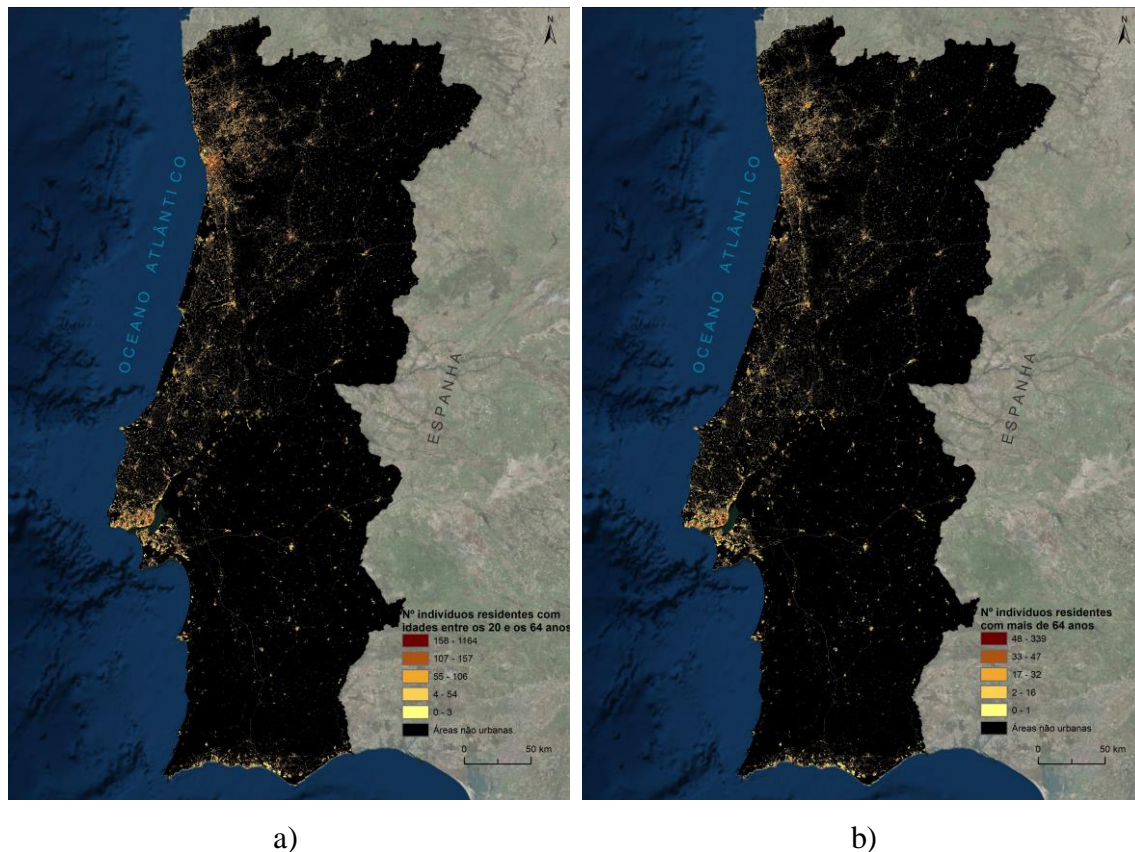


Figura 4.20: Distribuição do número de indivíduos residentes entre os 20 e os 64 anos (a) e com mais de 64 anos (b), em 2011

Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

Relativamente às famílias, estas podem ter 1, 2, 3, 4, 5 ou mais elementos, dependendo geralmente de fatores como a etnia, a situação profissional, o poder de compra, entre outros. Os inquéritos realizados pela Sonae revelaram que os agregados familiares dos visitantes das lojas Meu Super tendem a ser compostos por dois elementos. Assim sendo optou-se por estudar a distribuição das famílias clássicas compostas por 1 ou 2 elementos (Figura 4.21). Dos 5 945 860 de famílias clássicas existentes nas áreas urbanas em Portugal Continental, 2 042 032 são famílias compostas com 1 ou 2 elementos (34,3%). As áreas com maior aglomeração de famílias com 1 ou 2 elementos, à

semelhança com os indivíduos residentes, encontram-se ao longo do litoral com especial incidência em Lisboa e no Porto.

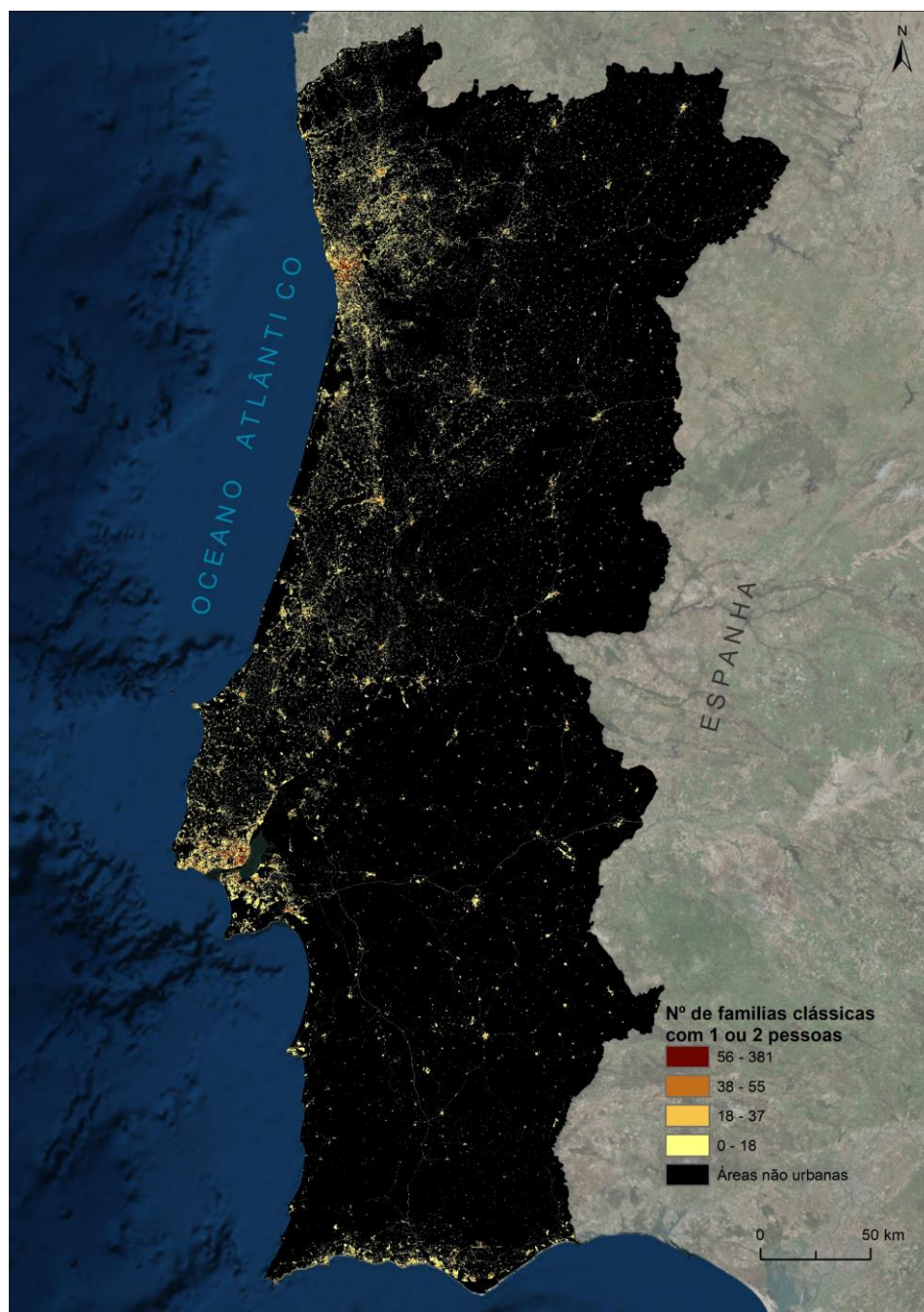


Figura 4.21: Distribuição do número de famílias clássicas compostas por 1 ou 2 elementos, em 2011

Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

Como referido anteriormente, para estudar o poder de compra da população recorreu-se a variáveis como a situação profissional, sectores de atividade da população e habilitações literárias.

A situação na profissão da população pode dar informações referentes ao poder de compra dos indivíduos, por exemplo, indivíduos empregados terão à sua disposição mais rendimento para gastar nas suas compras e/ou na deslocação, tendo, por isso, uma maior variedade de estabelecimentos à sua disposição. Um indivíduo pensionista ou reformado já terá menos escolhas à sua disposição uma vez que, como tendencialmente faz compras diariamente/semanalmente não estará disposto a percorrer longas distâncias para as fazer. Em Portugal, existem 4 109 976 indivíduos empregados e 2 223 335 indivíduos pensionistas ou reformados, sendo que as maiores concentrações das duas variáveis se encontram no litoral, com especial incidência nas cidades de Lisboa e Porto (Figura 4.22).

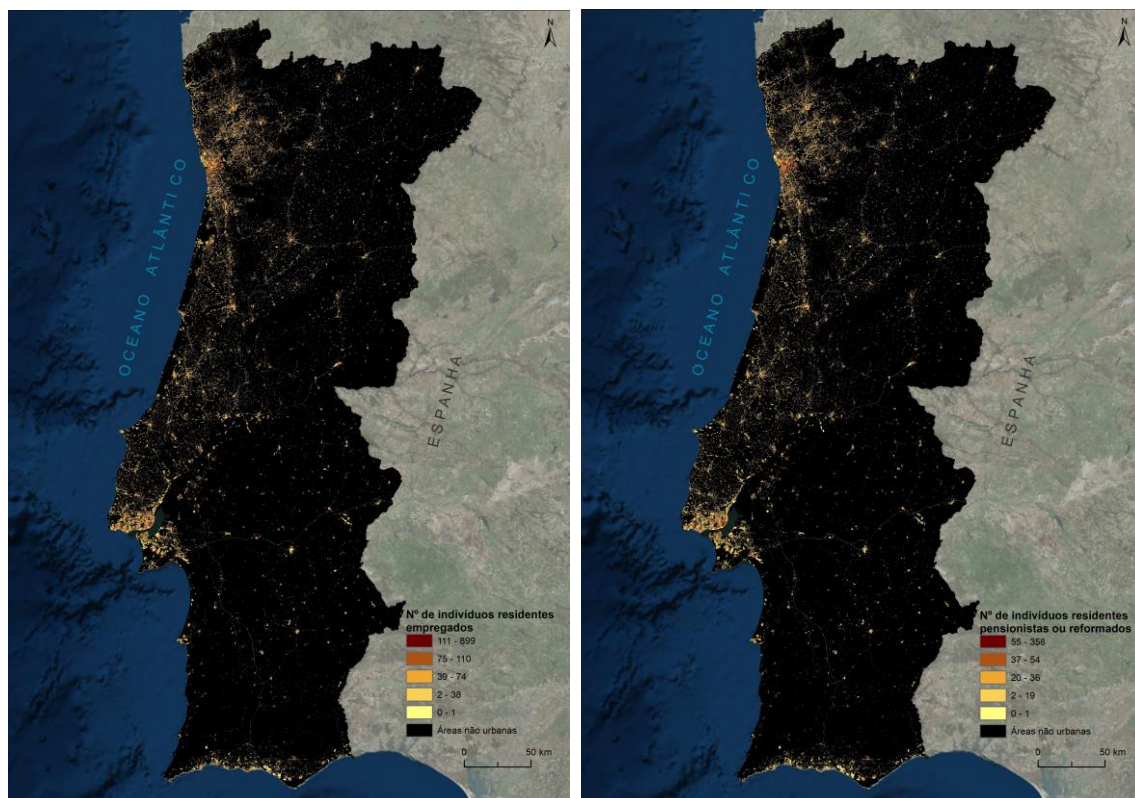
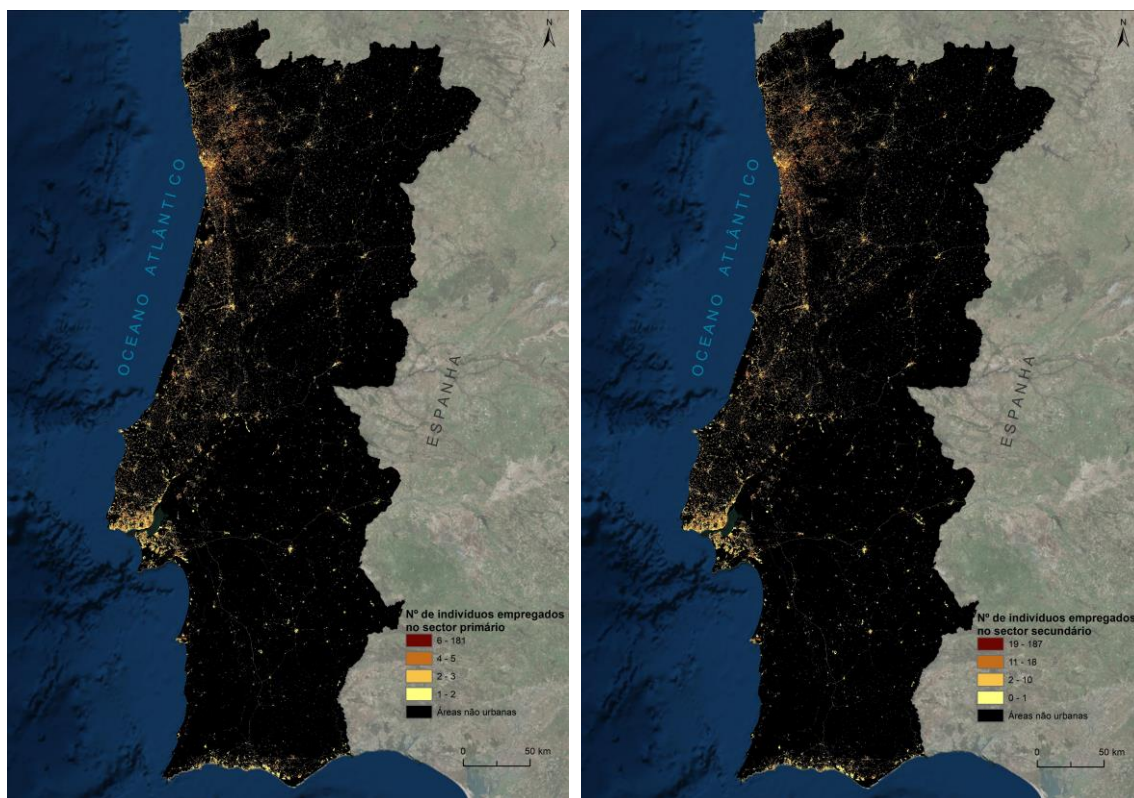


Figura 4.22: Distribuição do número de residentes empregados (a) e pensionistas ou reformados (b), em 2011

Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

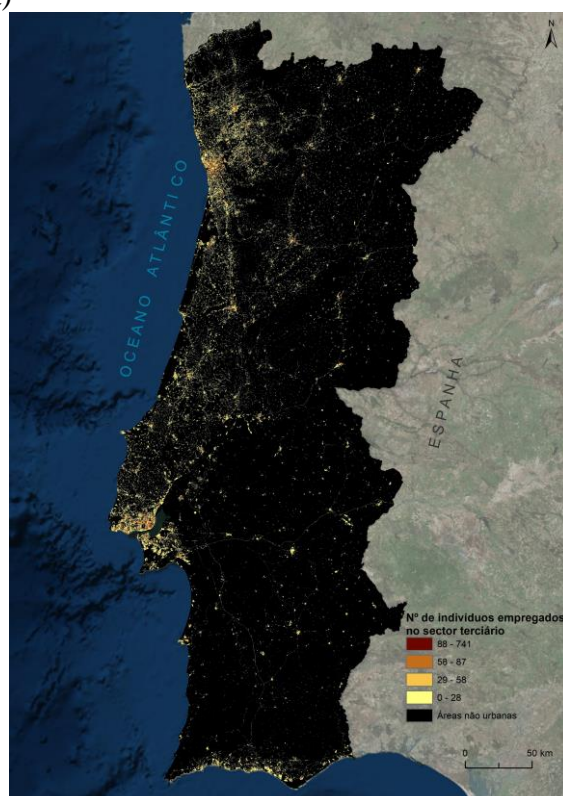
De modo a obter uma ideia mais aprofundada sobre o poder de compra da população, o número de residentes empregados por cada sector de atividade poderá informar-nos do rendimento auferido pela população empregada. As atividades incluídas no sector primário são a agricultura, pecuária, silvicultura, extração mineira, apicultura e pesca, ou seja, atividades que extraem recursos diretamente da Natureza, sem qualquer transformação. As atividades incluídas no sector secundário são a indústria, construção civil, obras públicas, fornecimento de água, gás e eletricidade, ou seja, atividades que transformam a matéria prima em produtos acabados ou semiacabados. As atividades incluídas no sector terciário são a educação, saúde, banca, seguros, transporte e turismo, ou seja, atividades que englobam o comércio e os serviços e inclui atividades que não produzem bens, mas prestam serviços. Por norma, os indivíduos empregados no sector primário auferem rendimento inferior quando comparados com os restantes sectores. Dos 4 109 976 indivíduos empregados, 2,8% estão empregados no sector primário (114 012 indivíduos), 26,8% no sector secundário (1 099 771 indivíduos) e os restantes 70,4% no sector terciário (2 891 199 indivíduos).

Relativamente à distribuição destas variáveis, assiste-se a uma maior concentração de indivíduos empregados no sector primário na região Norte (Figura 4.23a)), podendo dever-se à conjugação de fatores favoráveis à prática da agricultura e, consequentemente, a elevada presença de vinicultura na região. Em relação ao sector secundário (Figura 4.23 b)), este já se encontra mais distribuído, havendo ainda maior concentração na região Norte, mas com menos focos de concentração. Por último, a distribuição dos indivíduos empregados no sector terciário (Figura 4.23 c)) já se encontra a acompanhar a tendência nacional, havendo uma maior concentração dos serviços nas áreas metropolitanas de Lisboa e Porto.



a)

b)

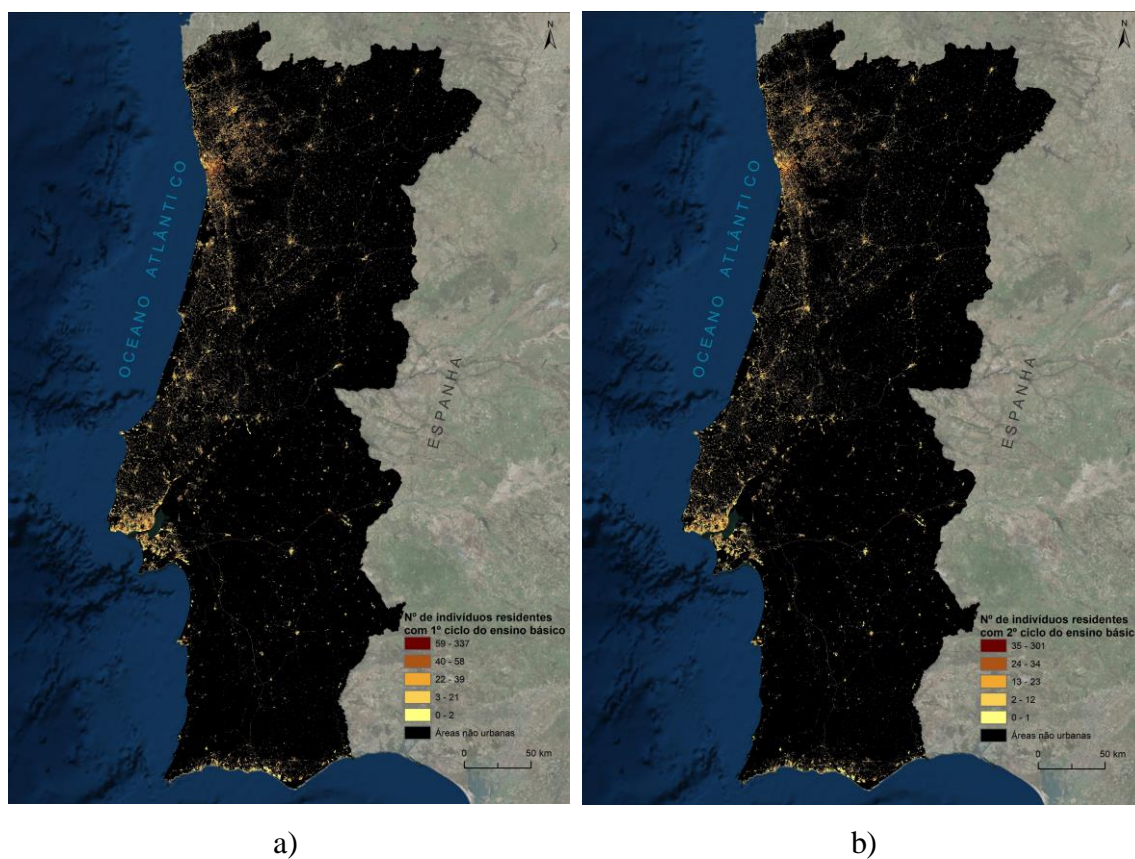


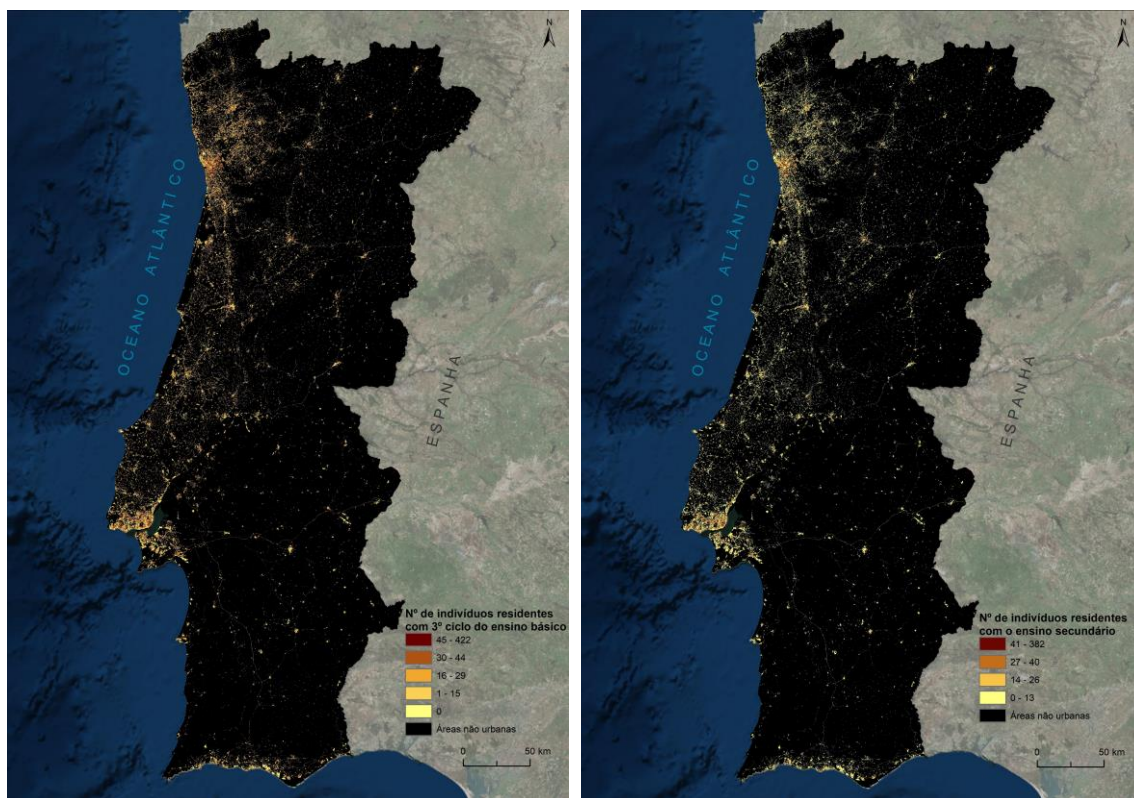
c)

Figura 4.23: Distribuição do nº de Indivíduos Residentes empregados no sector primário (a); N° Indivíduos Residentes empregados no sector secundário (b); N° Indivíduos Residentes empregados no sector terciário (c), em 2011

Base cartográfica: Caop (IGP), BGRI (INE), World Imagery (ESRI)

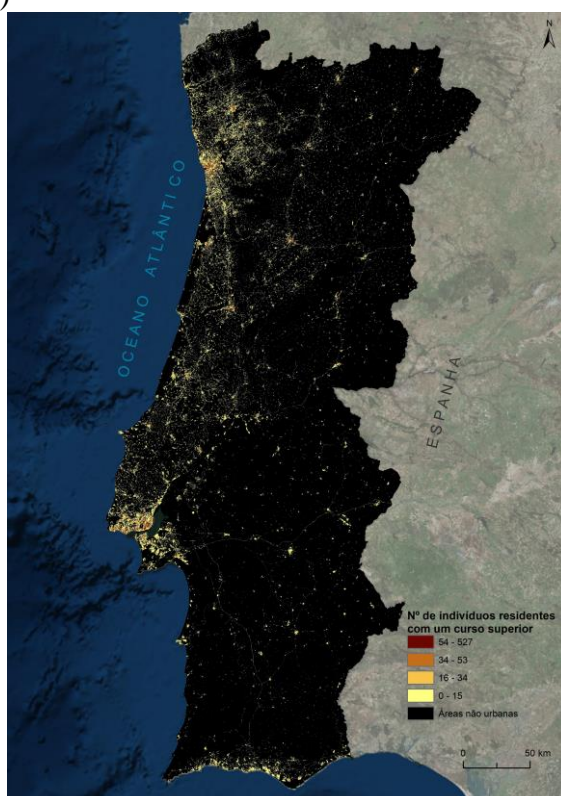
As habilitações literárias (Figura 4.24) da população também podem fornecer informações referentes ao poder de compra da população e, consequentemente, o local de compras da população. Indivíduos com o 1º ou 2º ciclo têm mais tendência a fazerem compras perto de casa uma vez que, por norma, são idosos, e apresentam rendimento baixo e, portanto, não fazem grandes deslocações para fazer as suas compras. Indivíduos com o ensino secundário (escolaridade obrigatória) e ensino superior, já apresentam, por norma, rendimento superior às pessoas com o 1º, 2º e 3º ciclo, e por isso já têm mais opções à sua disposição e mais informação ao seu dispor para escolherem a melhor opção para fazerem as suas compras.





c)

d)



e)

Figura 4.24: Distribuição do nº de Indivíduos Residentes com o 1º Ciclo (a); N° Indivíduos Residentes com o 2º Ciclo (b); N° Indivíduos Residentes com o 3º Ciclo (c); N° Indivíduos Residentes com o Ensino Secundário (d) e N° Indivíduos Residentes com um Curso Superior, , em 2011. Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

4.3. Infraestruturas

Além das variáveis sociodemográficas, existem outras variáveis que conseguem influenciar e explicar fenômenos. Neste subcapítulo, serão abordadas algumas variáveis utilizadas neste projeto, nomeadamente a densidade de estradas, a distância à concorrência, a densidade de lojas da concorrência, com o valor da área de loja, e a densidade de lojas Meu Super com o valor das vendas brutas, que poderão influenciar a escolha dos melhores locais para abrir novas lojas.

4.3.1. Estabelecimentos

Com a realização da análise geodemográfica na área de estudo, faltava obter as áreas onde a população não tem nenhuma oferta comercial ou onde a oferta de serviços comerciais é fraca. Nesta fase, a análise da densidade de *Kernel* pode identificar as áreas com maiores concentrações de potenciais clientes. A estimativa da densidade de *Kernel* é uma forma não-paramétrica de estimar a função de densidade de probabilidade de uma variável aleatória (Rosenblatt, 1956, citado por Roig-Tierno et al., 2013).

Conceptualmente, o objetivo da estimativa da densidade de *Kernel* é calcular a densidade dos pontos numa determinada área usando a distância entre os pontos, se e somente se os pontos tiverem o mesmo peso. No entanto, diferentes pesos podem ser atribuídos a cada ponto para atribuir maior importância aos pontos específicos em relação ao resto. É criada uma superfície suavemente curvada moldada a cada ponto, onde o valor mais elevado da superfície se encontra na localização do ponto e diminui com o aumento da distância ao ponto, atingindo zero na distância do raio de pesquisa do ponto (Silverman, 1986). Embora hajam diferentes tipos de modelos que usam estimativas de *Kernel*, Härdle (1991), citado por Roig-Tierno, Baviera-Puig, & Buitrago-Vera (2013), afirma que a escolha do modelo é quase irrelevante para a qualidade da estimação e, consequentemente, para o resultado da análise. Assim sendo, a função quadrática de *Kernel*, descrita por Silverman (1986) é das mais usadas.

Geralmente, e para o caso univariante, o estimador pode ser escrito da seguinte maneira:

$$f(x) = \frac{1}{nh} + \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (4.3)$$

Em que:

$f(x)$ = função da densidade de *kernel*

x = ponto onde a densidade é estimada

X_i = valor da variável no caso de $i=1, \dots, n$.

h representa o raio de pesquisa ou parâmetro de suavização. Este parâmetro limita a influência de cada dado a um campo, definido pelo raio de pesquisa. Quando o raio de pesquisa é maior, este causa um aumento da suavização no mapa resultante.

K = função *Kernel*.

Já a função de *Kernel* baseada na função quadrática do *Kernel*, descrita em Silverman (1986), é definida por

$$K = \frac{3}{\pi} (1 - u^2)^2, \text{ para } |u| \leq 1, \text{ onde } u = (x - h_i) / h \quad (4.4)$$

A função quadrática de *Kernel*, especificamente utilizada para estudos realizados em estrutura matricial, é caracterizada por um ambiente circular que é definido para cada píxel no mapa e, em seguida, é usado como uma linha de base. O centroide de cada píxel é o centro do círculo e os pontos que estão contidos são usados para formar o dividendo. Cada ponto é ponderado de forma diferente, de acordo com sua proximidade com o centroide do píxel: os pixéis mais próximos do centroide têm um peso maior do que aqueles mais distantes. A expressão utilizada é

$$L_j = \sum_{i \in C_j} \frac{3}{\pi r^2} \left(1 - \frac{d_{ij}^2}{r^2} \right) \quad (4.5)$$

Onde:

L_j é a densidade estimada para o píxel;

d_{ij}^2 é a distância entre os pontos i e j ;

r^2 é a largura do raio de pesquisa, que determina o grau de suavização;

$C_j = \{i \mid d_{ij} < r\}$, de modo a que o conjunto seja formado pelos pontos i cuja distância para o centroide da célula j seja menor que o raio do círculo prescrito.

A densidade de cada célula é calculada adicionando os valores de todas as superfícies de Kernel onde estas se sobrepõem ao centro da célula. O resultado é expresso em unidades de um fenómeno particular por unidade de área superficial.

Tendo em conta o objetivo, tornou-se imperativo conhecer como o mercado se distribuí. É necessário saber onde existem mais estabelecimentos, quer das lojas da insígnia em estudo, quer das lojas da concorrência. Para isso, recorreu-se à análise da densidade de lojas Meu Super, com base no número de vendas brutas efetuadas, e à densidade de lojas da concorrência, de acordo com a área das lojas.

4.3.1.1. Densidade de lojas Meu Super

Já é conhecida a localização de todas as lojas Meu Super, em funcionamento, no entanto, quando se pretende abrir uma nova loja, algumas lojas têm mais impacto que outras na população que abastece. Assim sendo, tornou-se importante conhecer as áreas onde as lojas Meu Super estão mais representadas e têm mais vendas brutas (Figura 4.25).

A Figura 4.25 mostra as áreas onde existe maior densidade de lojas com mais vendas brutas. Como se esperava as áreas onde não existem lojas Meu Super ou onde as vendas brutas são mais reduzidas, estão inseridas na primeira classe, até 368 unidades monetárias por km², representando, aproximadamente, 97% do território. A segunda classe, das 309 a 810 unidades monetárias por km², são as áreas com menor representação na área de estudo, representando apenas, aproximadamente 1%, uma vez que apenas são delimitadas estas áreas onde existe forte densidade de lojas com elevadas vendas brutas. Com apenas 2% da área em estudo, estão as áreas onde estão inseridas as lojas Meu Super e onde há mais concentração destas.

Poderá haver duas hipóteses possíveis para a localização das novas lojas: a primeira, em que as novas lojas poderão aproveitar a proximidade das lojas MS com mais vendas brutas por km², podendo assim dividir lucros entre duas lojas; ou, a segunda, onde a localização das novas lojas deverá ser nas áreas com menos vendas por km², podendo abrir novos estabelecimentos que poderão atrair clientes de outras áreas, onde não há

presença de lojas MS, e reduzir a probabilidade de canibalização das lojas Meu Super. Para esta dissertação, decidiu-se utilizar a segunda opção para não ocorrer a canibalização das lojas da insígnia.

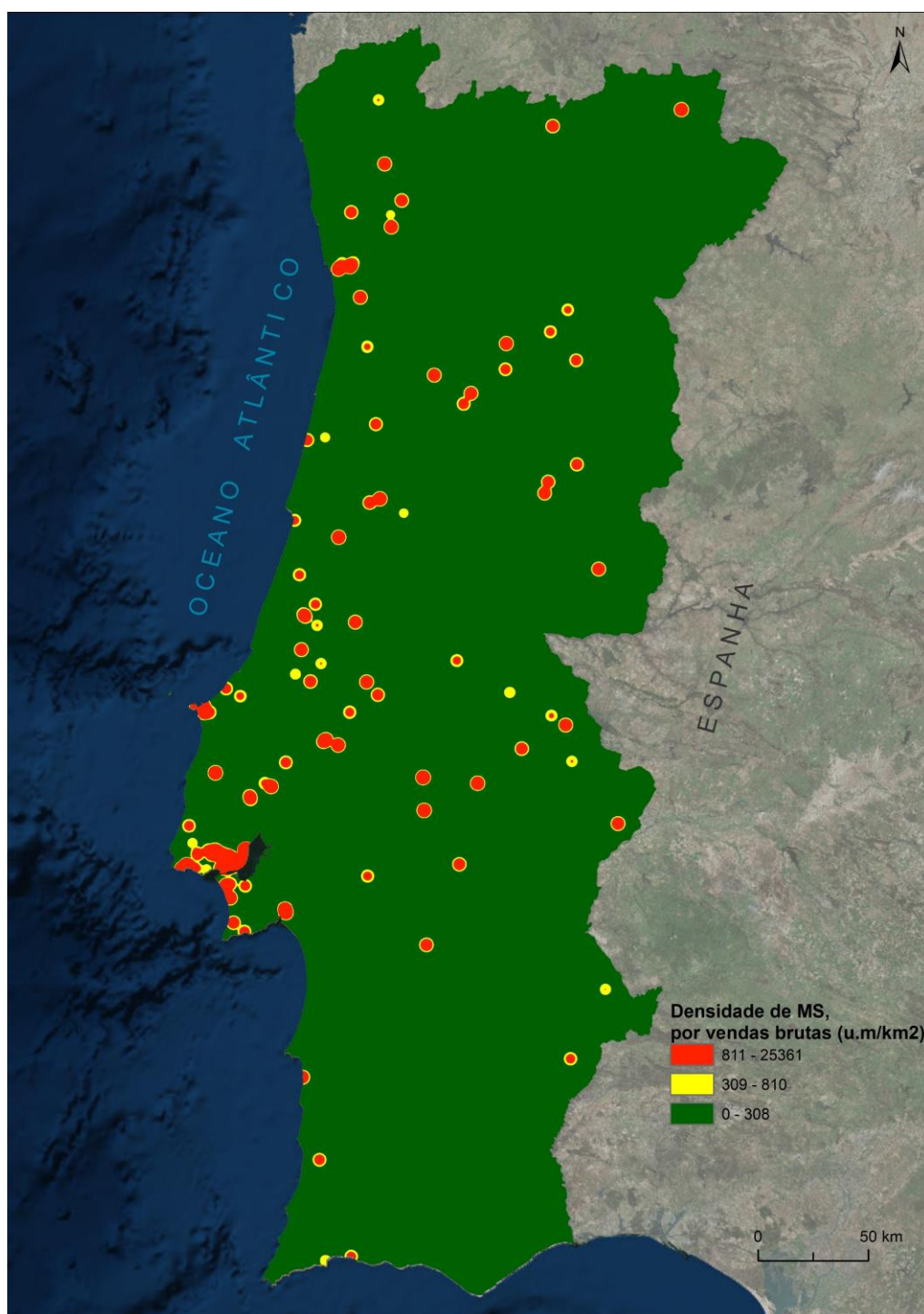


Figura 4.25: Densidade de lojas MS, por vendas brutas

Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

4.3.1.2. Densidade de lojas da concorrência

A concorrência é um dos principais fatores a serem considerados ao determinar a localização de uma nova loja. Por esta razão, é importante realizar uma análise detalhada para determinar a geocompetição na área de estudo de forma quantitativa e visual. Áreas com um maior espaço de retalho (áreas saturadas ou ocupadas), podem, deste modo, ser detetados. Da mesma forma, também são determinadas as áreas com menor espaço comercial ou com baixa oferta comercial. Conforme representado na Figura 4.26, há uma alta saturação em algumas capitais de distrito, pois é onde a maioria das áreas comerciais de diferentes lojas se sobrepõem.

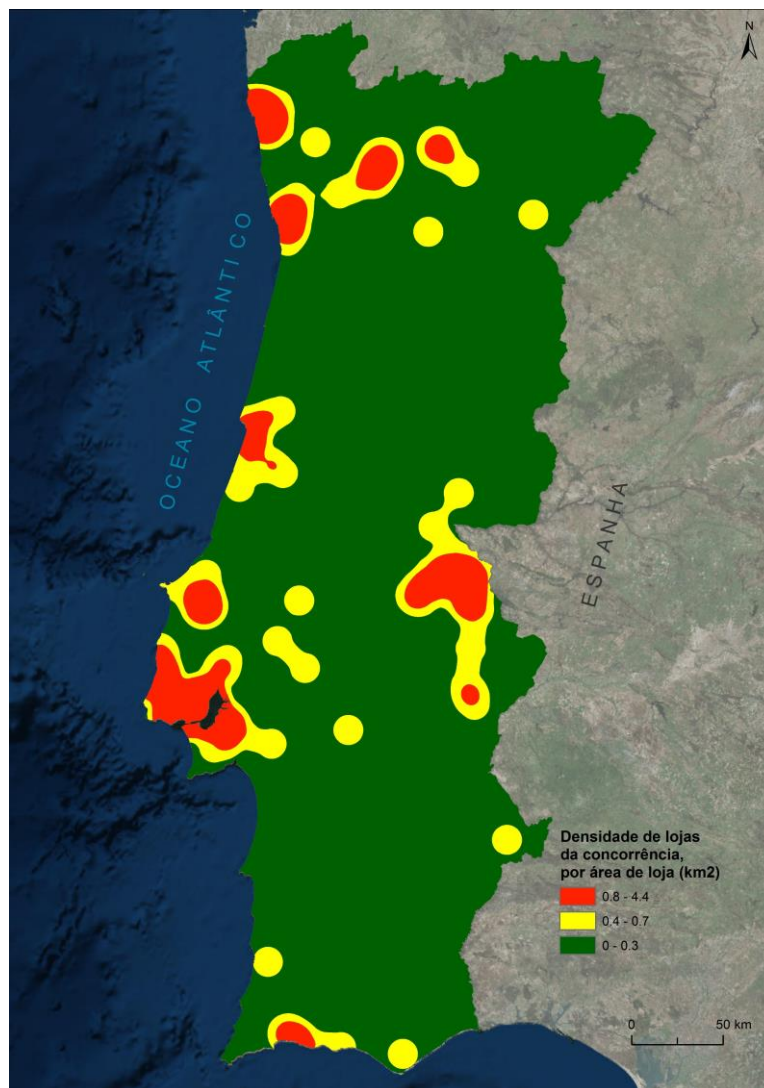


Figura 4.26: Densidade de lojas de insígnias concorrentes, por área de loja

Base cartográfica: CAOP (IGP), BGRI (INE), World Imagery (ESRI)

As áreas com maior presença de lojas da concorrência, localizam-se, maioritariamente nas grandes cidades no litoral e onde, normalmente, existe mais população, representando, aproximadamente, 6% do território. Existe também, no interior, em Évora, uma grande densidade que poderá ser explicada pela dimensão das lojas aí presentes. Às áreas onde existem lojas muito dispersas foi atribuída a menor classificação, representando cerca de 85% do território.

4.3.2. Densidade de estradas

Para além do já referido, existem mais variáveis que influenciam tanto a localização da oferta como a da procura, sendo a rede viária uma das mais importantes. Esta torna-se um fator determinante para a localização de estabelecimentos comerciais, visto que influencia a fixação da população e melhora a acessibilidade a determinados locais. Sendo o objetivo a localização de novas lojas, é indispensável que as zonas para onde se pretende expandir tenham boa acessibilidade. Para isso, tornou-se necessário uma variável que avaliasse a acessibilidade da população. Através da densidade de estradas, criada através da ferramenta *Line Density* (densidade de linhas), foi possível determinar as áreas onde existem mais estradas e, consequentemente, onde existe maior acessibilidade.

A ferramenta *Line Density* calcula a densidade de recursos lineares na vizinhança de cada célula de *output*. A densidade é calculada em unidades de comprimento por unidade de área. Conceitualmente, um círculo é desenhado em torno de cada centro das células matriciais, usando o raio de procura. O comprimento da porção de cada linha que se enquadra no círculo é multiplicado pelo valor do campo População. Estes resultados são somados, e o total é dividido pela área do círculo (ESRI, sem data-c). A Figura 4.27 é o resultado deste processo.

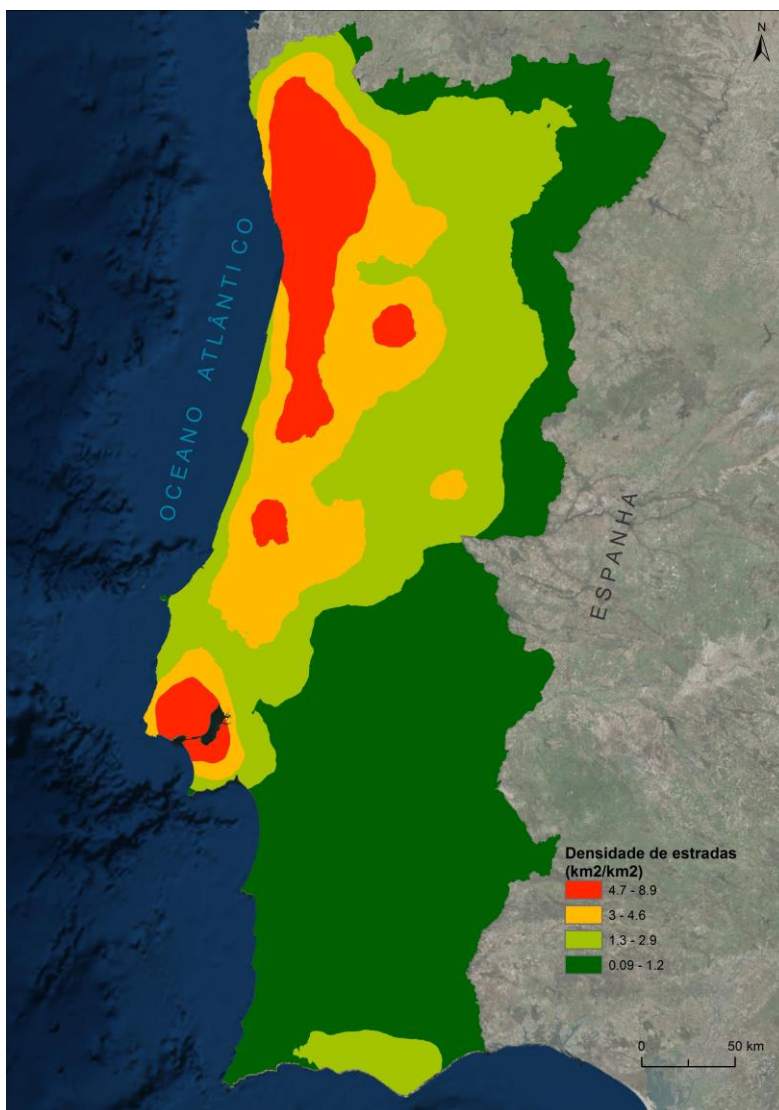


Figura 4.27: Densidade de estradas

Base cartográfica: CAOP (IGP), Rede Viária (NOKIA), World Imagery (ESRI)

Como seria de esperar, é o interior e sul do país que apresenta menor densidade de estradas. Esta classe representa cerca de 47% do território, uma vez que é nestas áreas onde há menor aglomeração de estradas, assim como menos população. Por outro lado, as áreas com maior densidade de estradas são as áreas metropolitanas de Lisboa e Porto e algumas áreas ao longo do litoral norte e centro, representando cerca de 9% da área em estudo.

4.3.3. Distância à concorrência

Outro fator a ter em consideração é a distância das lojas Meu Super aos seus concorrentes diretos. Esta variável poderá informar-nos sobre que lojas concorrentes das lojas Meu Super estão mais próximas, podendo definir um padrão comum por todo o território. Para a criação desta variável foi utilizada a ferramenta *Euclidean Distance* para calcular a distância em linha reta entre as lojas em estudo e as concorrentes. O resultado que se obteve foi uma tabela com a distância de cada loja MS à loja concorrente mais próxima dessa loja.

4.4. Áreas de Influência

As áreas de influência não serão uma variável *per si*, no entanto, estas vão indicar qual a importância, em termos de amplitude espacial, que algumas das variáveis acima descritas possuem, permitindo observar a correlação entre as variáveis que será aprofundada no próximo capítulo.

De acordo com a *American Marketing Association* (AMA), a área de influência é a área geográfica onde estão inseridos os consumidores de uma determinada empresa particular ou grupo de empresas para a comercialização de bens ou serviços específicos (Berman & Evans, 1995). Trata-se de um conceito essencial nos estudos de localização, uma vez que reflete a distância que os clientes estão dispostos a percorrer para a compra de determinado bem ou serviço.

As áreas de influência têm sido bastante estudadas ao longo dos tempos, devido à sua importância no desempenho de várias empresas, com especial foco no comércio a retalho, tendo despertado interesse por parte de autores como Reilly (1929), Huff (1964) e Applebaum (1966), e ganho renovado interesse com a introdução dos SIG, visto que estas tecnologias permitem integrar e processar grandes volumes de dados, e permitem relacionar dados que não seria possível de outro modo, originando mapas com áreas de influência que facilitam as análises de mercado.

O desempenho dos supermercados depende da localização, uma vez que a maior parte das vendas de uma loja provêm de clientes que residem dentro de uma área geográfica relativamente pequena em redor da loja (Parente & Kato, 2001). Applebaum (1966, p.127), identificou a área de influência de supermercados através de *customer spotting*, isto é, através de pesquisas dos consumidores nas lojas já existentes, no local

onde residem, para determinar a área de influência primária de um conjunto de lojas através do mapeamento de clientes. Com o cruzamento desta informação com as vendas reais das lojas, tem-se base para avaliar as vendas de localizações futuras. Esta técnica consiste em identificar num mapa a localização dos clientes de uma loja e, mediante a sua dispersão, o autor identificou três segmentos de uma área de influência:

- Área de influência primária: região mais próxima da loja, que apresenta maior densidade de clientes, onde estão concentrados cerca de 60% a 75% dos clientes;
- Área de influência secundária: região em redor da área de influência primária, onde estão cerca de 15% a 25% dos clientes;
- Área de influência terciária: região que contém a porção restante dos clientes que residem mais afastados da loja (cerca de 10%).

Ou seja, a área de influência compreende a extensão máxima que os clientes estão dispostos a percorrer para efetuarem as suas compras. Assim, quanto mais próximo estiver um negócio do seu público-alvo, maior será a sua área de influência na região e o atendimento aos clientes mais personalizado, o que poderá gerar um maior grau de fidelização e publicidade boca-em-boca.

Além destas classificações de áreas de influência, ainda existem dois tipos de representação das áreas de influência. Estas podem ser representadas por isócronas ou isocotas. As primeiras delimitam as áreas de influência de acordo com o tempo, enquanto as segundas as delimitam por distância. De modo a obter uma representação com uma maior aproximação da realidade, nesta dissertação, optou-se por delimitar as áreas de influência através de medições de rede, para isso foi necessária a sua construção.

4.4.1. Construção da rede

Uma rede é um conjunto de linhas interconectadas que permitem representar a circulação ou movimento de um transporte, pessoas, bens comunicações, entre outros (González, 2012, citado por Figueiredo, 2014). Ou seja, é um conjunto de pontos ligados por linhas em que estabelecem várias relações entre si, de acordo com as características inerentes às linhas.

O estudo das redes provém de uma subdisciplina da Matemática denominada “Teoria dos Grafos”, que efetivamente estuda a topologia das redes (Kemp, 2008). O

grafo representa graficamente o fenómeno num diagrama (conceito primitivo), constituído por nós (*e.g.* cruzamentos, pontos de origem ou de partida), que correspondem aos pontos, e por arcos (*e.g.* ruas, pontes) que correspondem às linhas. Segundo Figueiredo (2014), os grafos são geralmente representados de forma sintética e temática, respeitando a topologia e face à distância real entre os pontos e a sua posição real, relativizando as localizações com o intuito de somente mostrar graficamente as ligações decorrentes (*e.g.* estações de metropolitano). A conectividade de uma rede é dada pelo número de ligações existentes entre os nós, isto é, quantos mais arcos estiverem associados a um maior número de nós, maior será a conectividade da rede. Deste modo, é possível proceder a análises sobre as medidas de conectividade e de ligação, de modo a analisar a estrutura da rede (Haggett & Chorley, 1969).

Nesta dissertação foi elaborada uma rede através de uma extensão e módulo de construção e análise de redes, o *Network Analyst*, a partir de um *software* de SIG, o ArcGIS, utilizando a estrutura topológica *Network Dataset* para a construção de topologias de linhas.

As *Network Dataset* são um conjunto de dados SIG projetados para suportar a análise da rede. Geralmente consistem em linhas que representam as rotas de fluxo na rede, enriquecidas com outros recursos (como pontos de junção), topologia e atributos que modelam propriedades relevantes para a rede, como impedância e capacidade de fluxo. As *Network Dataset* são o conjunto de dados mais adequado para modelar redes de transporte, uma vez que estas são criadas através de recursos simples (linhas (segmentos) e pontos (nós)) e arcos, que permitem armazenar a conectividade dos recursos (ESRI, sem data-e). A geometria destes recursos permite a conectividade da rede. Estes elementos consistem em (ESRI, sem data-d):

- Arestas (*edges*): que consistem nas ligações onde os agentes viajam, e são conectados aos outros elementos, nomeadamente as junções (*junctions*);
- Junções: que conectam as arestas e permitem a conectividade entre elas;
- Arcos (*turns*): que armazenam informações que podem afetar o movimento entre duas ou mais arestas.

Especificamente para esta dissertação, criou-se um atributo na variável da rede viária, onde se calculou o tempo da deslocação a pé, partindo-se do princípio que uma pessoa, em uma hora (60 minutos), percorre 3000 metros (em anexo). A velocidade a que

um peão se desloca é influenciada pelas suas características intrínsecas, como a idade, a condição física ou as suas limitações. O valor referido foi escolhido de acordo com a ideia de que é a população idosa que mais utiliza este tipo de loja, mas não só, e estes por norma movimentam-se mais lentamente. Para as vias onde não é possível percorrer a pé, foi impossibilitada essa opção com a atribuição do valor ‘-1’. Através da criação destes atributos foi possível a conceptualização da rede através da criação de uma *Network Analyst*, que permitiu, posteriormente, analisar a rede.

A extensão do *Network Analyst* ajuda a realizar análises espaciais com base na análise dessas redes, possibilitando criar rotas multimodais, procurar o ponto mais próximo, criar áreas de influência ou calcular custos matriciais de origem-custo (ESRI, 2013, citado por Figueiredo, 2014). Estas análises apresentam um conteúdo tanto mais realista quanto maior o detalhe e veracidade da informação inserida, associada às condições da rede, nomeadamente, às respetivas variáveis, atributos ou restrições. Com a delimitação das áreas de influência, ou áreas de serviço (designação da ESRI), é possível definir polígonos, com áreas de atuação, em função de um fator ou relação de vários fatores inerentes à rede (*e.g.* tempo, velocidade, distância). A cada elemento ou entidade da rede (nós ou arcos) são armazenados numa base relacional, atribuindo a cada nó ou arco um identificador (ID) e em que cada arco vai estar associado ao nó de origem e de destino dando o sentido ou direção ao arco (Goodchild, 2003).

Para esta dissertação a sua aplicação permitirá delimitar e analisar áreas de influência (ou superfície de custo), a partir do atributo de custo tempo. A questão que se impunha era qual seria o período mais indicado para as áreas de influência dos estabelecimentos de proximidade. Burton & Mitchell (2006), afirmaram que as áreas de influência das lojas de proximidade são de 500 metros, no entanto, Cachinho et. al, citados por Afonso (2018), afirmam que o tempo máximo que um consumidor está disposto a percorrer para fazer as suas compras diárias e para transportá-las é 10 minutos. Dadas as diferentes “opiniões” optou-se por analisar três tipos de áreas de influência, nomeadamente a 5, 7 e 10 minutos das lojas das insígnias em estudo (Figura 4.28).

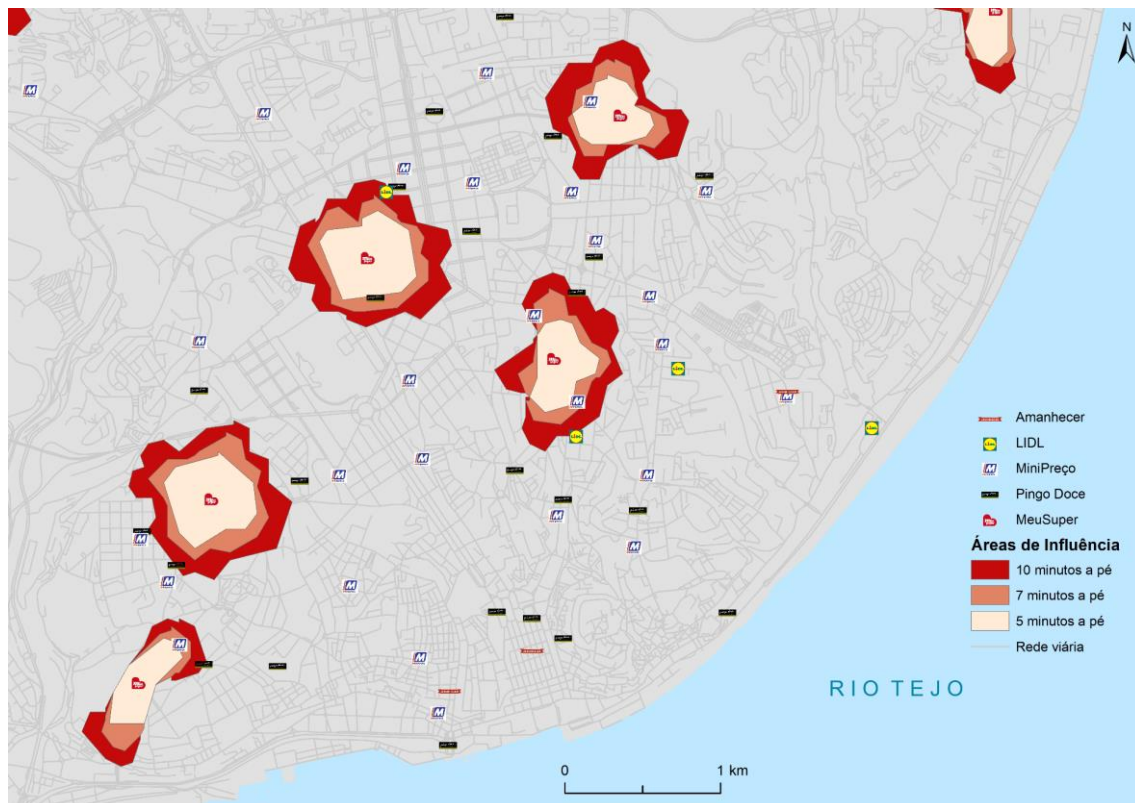


Figura 4.28: Exemplo das áreas de influência das lojas Meu Super

Base cartográfica: CAOP (IGP), Rede Viária (Nokia)

Após a análise das áreas de influência, determinou-se que a área que corresponde mais realisticamente à das lojas Meu Super são as áreas a 5 minutos, uma vez que é nestas áreas em que a área de influência apenas tem a insígnia em estudo ou é partilhada apenas por uma outra insígnia. Estas áreas também permitem a ida à loja em 5 minutos e voltar a casa em 5 minutos (10 minutos como definidos por Afonso (2018)) e menos de 500 metros, de ida e volta da loja, definidos por Burton & Mitchell (2006).

5. Location Analytics

Location Analytics consiste no processo ou na capacidade de obter informações sobre a localização ou a componente geográfica dos dados de negócios. Os dados, especialmente os gerados pelas empresas, contêm uma componente geográfica que, quando estabelecido num sistema de informação geográfica, permite obter novas dimensões de análise e perspectivas, neste caso, através de uma abordagem mais visual. Há uma percepção cada vez maior de que, ao adicionar localização geográfica aos dados empresariais e mapeá-los, as organizações podem melhorar drasticamente as suas ideias de dados tabulares. Mapas e análises espaciais fornecem um contexto totalmente novo que simplesmente não é possível com tabelas e gráficos (ESRI, 2012).

As empresas de retalho podem beneficiar fortemente deste tipo de análise (locativa) e em conhecer os hábitos de gastos de pessoas de diferentes locais geográficos ao longo do tempo, visto que poderão conseguir aumentar as vendas e aumentar a retenção de clientes através de anúncios direcionados adequados, e uma melhor distribuição dos produtos.

Neste capítulo encontrar-se-á o foco desta dissertação, visto que será aqui onde se fará o pré-processamento dos dados (lógica *fuzzy* e *Data Mining*); onde se criará modelos estatísticos para determinar os melhores locais para a abertura de novas lojas, e validar-se-á os resultados; e ainda, numa última fase se irá recorrer à análise de redes, mais concretamente, a modelos de localização-alocação para determinar as melhores localizações para as lojas.

5.1. Data Mining

Vivendo na era da informação, e com a onnipresença dos computadores, é-nos mais fácil encontrar informação que anteriormente era desaproveitada. Atualmente, todas as ações praticadas, seja no supermercado, nos bancos ou na *World Wide Web* (WWW), são armazenados, ou seja, a quantidade de informação tem crescido em dimensão e complexidade, tornando-se necessária a procura de padrões nos dados de forma a compreendê-los. Desde sempre se tenta encontrar padrões, quer na Natureza quer em dados eletrónicos. Assim, técnicas de *Data Mining* tornam-se uma ferramenta importante na procura de padrões de distribuição espacial.

As técnicas de *Data Mining* são uma das etapas do processo de Extração de Conhecimento em Bases de Dados (ou *Knowledge Discovery in Databases*). Este processo consiste num processo iterativo de extrair informação válida, previamente desconhecida e de máxima abrangência a partir de grandes bases de dados, de modo a melhorar a compreensão de um conjunto de dados e melhorar a tomada de decisão. Este é constituído por cinco etapas básicas: seleção, pré-processamento, transformação, *Data Mining* e avaliação/interpretação (Figura 5.1).

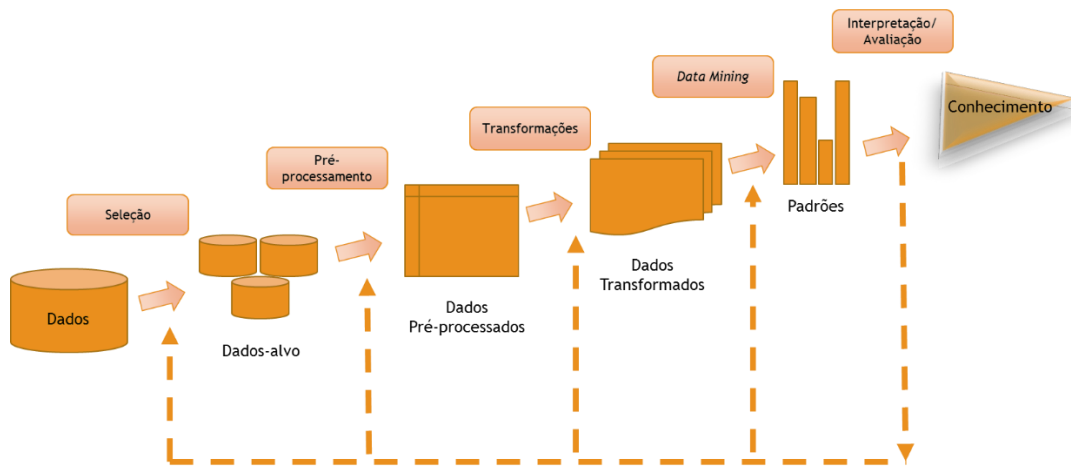


Figura 5.1: Etapas do processo de Extração de Conhecimento em Bases de Dados

Fonte: Fayyad et al. (1996, p.41)

Qualquer processo de ECBD começa com a aquisição de uma grande quantidade de dados que terá de passar por um processo de seleção de dados para ficar apenas com os dados que são relevantes para o trabalho, isto vai demonstrar um domínio do tema e dos dados empregues. A segunda e terceira fase passam pelo pré-processamento e limpeza dos dados e a transformação dos dados (reduções ou projeção dos dados em função do objetivo de *Data Mining*). É nestas primeiras etapas onde se consome cerca de 60% a 80% do tempo e esforço empregue no trabalho. A etapa intitulada de *Data Mining* é determinada pela aplicação de algoritmos específicos para a extração de padrões dos dados, podendo optar-se por uma das operações de DM (classificação, regressão, segmentação, associação ou sintetização), em função do objetivo do modelo a ser derivado. A última fase, como o nome indica, consiste na interpretação e avaliação dos padrões encontrados através das técnicas de DM.

De acordo com Berry e Linoff (1996), o objetivo do processo de DM é fornecer às empresas informações que possibilitam construir as melhores estratégias de marketing, vendas e suporte, melhorando assim os seus negócios. Atualmente, os investigadores e técnicos de marketing dependem de SIG modernos, e outras aplicações de cartografia assistida por computador como auxílio para as suas análises. Os SIG (ou mais especificamente, um Sistema de Apoio à Decisão Espacial (SADE)) oferecem melhores técnicas de visualização geográfica, o que auxilia na compreensão e capacidade de tomada de decisão, tornando-as melhores, mais rápidas e robustas (Yasnoff & Miller, 2014). Assim, *Data Mining* pode ser definido como a análise de grandes quantidades de dados e extração de informação desses dados, com o intuito de encontrar relações e sumariar os dados de forma a serem compreensíveis e úteis ao analista (Hand et al., 2001). Esta etapa preenche lacunas da estatística aplicada e da inteligência artificial na gestão de base de dados, organizando a informação de forma eficiente, permitindo a aplicação a grandes conjuntos de dados.

Os principais objetivos de *Data Mining* são a previsão e a descrição, sendo, deste modo, possível diferenciar as várias técnicas de DM, em modelação descritiva e preditiva, respetivamente. A primeira refere-se à descoberta de padrões descritivos dos dados que sejam interpretáveis; a modelação preditiva envolve a utilização de variáveis ou campos na base de dados para prever valores desconhecidos ou futuros de outras variáveis de interesse. Apesar de esta distinção não ser rígida, uma vez que alguns modelos preditivos podem ser descritivos, dado que é possível interpretá-los e vice-versa, com esta distinção é possível esclarecer qual o objetivo da técnica de DM a realizar (Fayyad et al., 1996).

Os objetivos destes tipos de modelação podem ser alcançados usando uma variedade de métodos/tarefas de DM, entre os quais:

❖ Modelos Descritivos (Fayyad et al., 1996, p.44-45)

- Segmentação (*Clustering*), que consiste na procura de um número finito de agrupamentos (*clusters*), exclusivos e exaustivos, hierárquicos ou sobrepostos, para descrever os dados. Ou seja, este método consiste divisão de uma população heterogénea em subgrupos homogéneos, de acordo com a semelhança entre eles, não havendo classes predefinidas;

- Sintetização, que envolve métodos para encontrar uma descrição compacta para um subconjunto de dados, podendo para isso fazer-se uma tabela com os valores de média e desvio-padrão para todos os campos dos dados;
- Modelação de dependências, que consiste em encontrar um modelo que descreva as dependências entre as variáveis;

❖ Modelos Preditivos

- Classificação, que consiste na construção de um modelo para ser aplicado a dados não classificados (Berry & Linoff, 1996), categorizando-os num conjunto predefinido de classes;
- Estimativa ou Regressão, enquanto a classificação trabalha com valores discretos, a estimativa trata valores contínuos, fazendo corresponder determinados intervalos de valores contínuos a diferentes classes.

De forma a desenvolver o modelo de localização, com base nos dados tratados anteriormente, recorreu-se ao coeficiente de correlação de Pearson para observar a relação entre as variáveis. Este coeficiente, que corresponde à tarefa de modelação de dependências, foi descrito por Karl Pearson, em 1897, e mede o grau de correlação linear (e a direção dessa correlação – se positiva ou negativa) entre duas variáveis, ou seja, o coeficiente de correlação de Pearson (r) é uma medida de associação linear entre variáveis (Britto et al, 2009).

Para entender este coeficiente é necessário entender primeiro dois conceitos-chave: associação e linearidade. Consideram-se duas variáveis estatisticamente associadas quando estas têm *scores* semelhantes, devido à distribuição das frequências ou pela partilha da variância, como é o caso da correlação de Pearson. Por outro lado, a linearidade implica que o aumento ou decréscimo de uma variável gera o mesmo impacto noutra variável (Britto et al, 2009).

Assim, este coeficiente pode ser executado da seguinte forma:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum (x_i - \bar{x})^2\right)\left(\sum (y_i - \bar{y})^2\right)}} \quad (5.1)$$

Em que x_i e y_i representam os valores das variáveis X e Y, e \bar{x} e \bar{y} são respetivamente as médias dos valores x_i e y_i . Se $r > 0$, x_i e y_i são positivamente correlacionados; se $r < 0$, x_i e y_i são negativamente correlacionados; se $r = 0$, x_i e y_i não dependem linearmente uma da outra.

Este coeficiente varia entre -1 e +1, onde -1 indica perfeita correlação negativa, e +1 perfeita correlação positiva, e em que 0 significa que não há correlação. Caso não exista correlação linear entre as variáveis pode existir correlação não-linear; existindo a necessidade de se avaliar cada situação e, caso seja necessário, de se investigar essa relação através métodos mais apropriados.

Assim, através do *Rstudio*, procedeu-se à aplicação do coeficiente de correlação de Pearson. Como resultado obteve-se a Figura 5.2, em que quanto mais forte é a cor azul e mais completo o gráfico circular, mais forte é a correlação positiva (mais perto de 1), por outro lado, quanto mais forte a cor vermelha e mais completo o gráfico circular, mais forte é a correlação negativa (mais perto de -1).

Através destes resultados, com o apoio da matriz de correlação gerada (em anexo), não foi possível, à primeira vista, determinar quais as melhores variáveis para utilizar nos modelos, uma vez que as variáveis provenientes dos censos apresentavam correlação positiva muito forte entre elas, como seria de esperar, e as variáveis criadas, anteriormente, apresentavam valores de correlação positiva baixa entre elas, e até valores de correlação negativa entre estas e as variáveis dos censos.

Quando duas variáveis apresentam elevado grau de correlação, estamos perante um caso de multicolinearidade, ou seja, as variáveis independentes apresentam relações lineares exatas ou aproximadamente exatas.

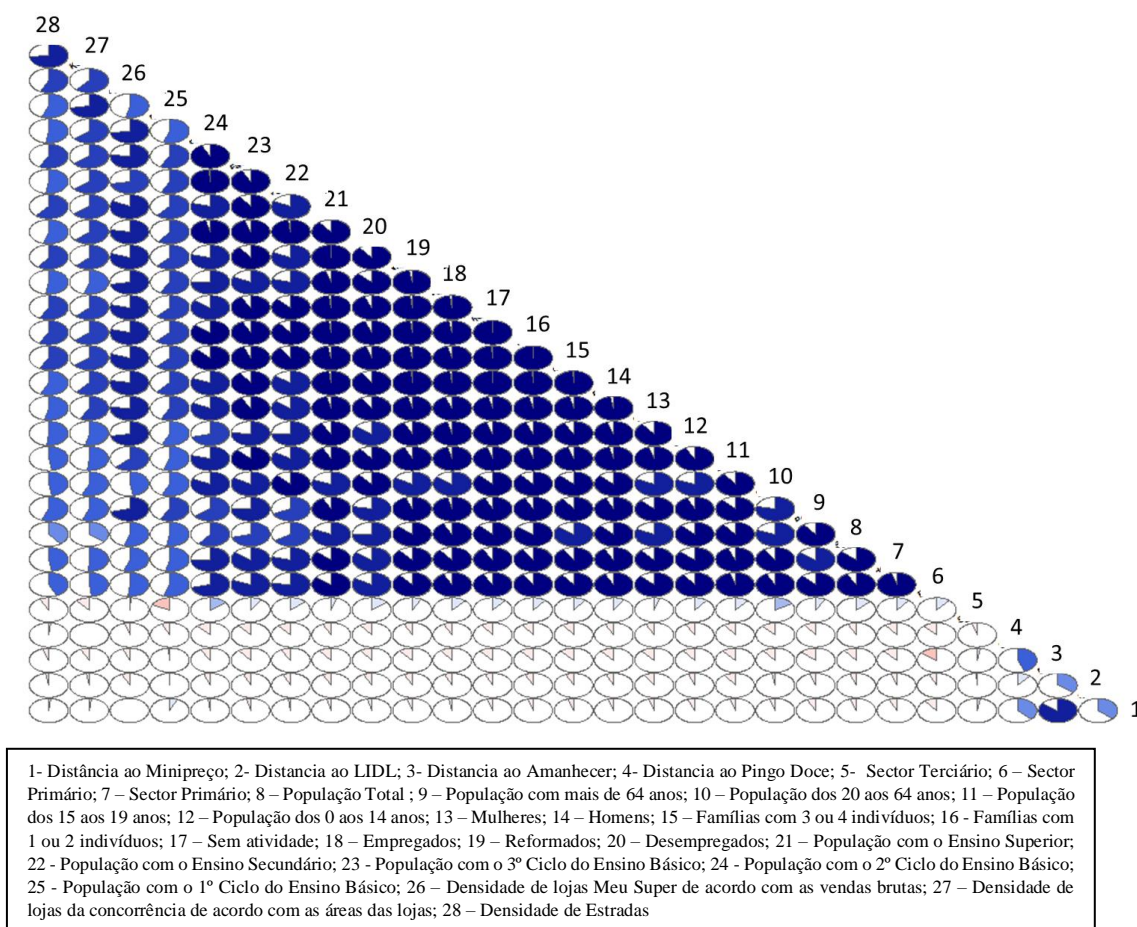


Figura 5.2: Correlação entre as variáveis

Assim, para escolher as melhores variáveis recorreu-se ao *Variance Inflation Factor* (VIF). Este fator fornece um índice que mede o quanto a variação (o quadrado do desvio-padrão da estimativa) de um coeficiente de regressão estimado aumenta devido à colinearidade entre as variáveis. Ou seja, mede o quanto a variação de um coeficiente de regressão estimado aumenta quando as variáveis estão correlacionadas. Quanto maior o valor, pior, visto que o objetivo é obter estimativas precisas. Se a variação dos coeficientes aumentar, o modelo não será tão confiável. O VIF é calculado para cada variável explicativa, da seguinte forma:

$$VIF = \frac{1}{1 - r^2} \quad (5.2)$$

Onde r^2 é o coeficiente de determinação da equação de regressão. O valor do VIF vai dar a magnitude da multicolinearidade (PVANB, 2016).

Descoberto o valor VIF para cada variável, escolheram-se as variáveis que apresentavam valores menores que 4, visto que para variáveis com VIF inferior a 4 é raro a presença de multicolinearidade. Regra geral, os valores de VIF que excedam 4 necessitam de investigação adicional, enquanto os valores de VIF excedendo 10 são sinais de multicolinearidade, e requerem correção (PennState Eberly College of Science, sem data). Visto isto, as variáveis escolhidas para a realização dos modelos são as referidas na Figura 5.3.



Figura 5.3: Variáveis escolhidas para os modelos (resultado do VIF)

5.2. Pré-processamento (Normalização das variáveis)

Os modelos são simplificações da realidade. Os modelos preditivos dependem de modelos estatísticos empíricos, que por sua vez dependem da natureza do conjunto de dados. O estudo de padrões espaciais, muitas vezes complexos e mal compreendidos, beneficia dessa abordagem amplamente independente de qualquer conhecimento pré-existente de processos subjacentes. Tal como os modelos estatísticos, os modelos difusos

(*fuzzy*) adaptam-se bem ao desenvolvimento de modelos empíricos, preditivos e *data driven* (Puig, 2011).

Com o rápido desenvolvimento dos Sistemas de Informação Geográfica, a forma de recolha, armazenamento, análise e disposição da informação alterou-se, conseguindo-se adaptar as informações a diferentes fenómenos. Por norma, diferentes critérios apresentam valores distintos, o que os torna incomparáveis entre si, inviabilizando a sua agregação imediata. Assim, de modo a resolver este problema, procede-se à normalização dos valores dos critérios, atribuindo a mesma escala de valores aos critérios, conseguindo minimizar os problemas oriundos do uso de diferentes escalas entre os critérios. Dependendo do tipo de fenómeno em estudo, este pode demonstrar um grau de indefinição ou incerteza (*fuzzy*), muitas vezes associado ao processo de tomada de decisão, que não poderá ser expresso com conjuntos nítidos de limites de classe (Kainz, 2001).

Na base da modelação cartográfica nos SIG está a lógica booleana. Esta é de natureza nítida (*crisp*), determinista e exata, não dando margem para imprecisões na informação geográfica e nos processos de tomada de decisões. Esta abordagem consiste na atribuição de apenas um atributo a cada entidade espacial, impondo limites nítidos, permitindo, simplesmente, que uma entidade espacial possa pertencer ou não a um conjunto (Sui, 1992). No entanto, esta técnica convencional de modelagem cartográfica provou ser de bastante difícil aplicação em algumas aplicações SIG, onde a imprecisão prevalece, porque nem todas as entidades na base de dados espacial podem ser definidas exclusivamente, seja no conjunto de atributos ou no seu delineamento espacial (Leung, 1988, citado por Sui, 1992).

Para fazer face a este tipo de problemas, surgiu a lógica *fuzzy*. Esta abordagem introduziu conceitos como "íngreme", "próximo" ou "adequado" (Kainz, 2001) para caracterizar os critérios de um fenómeno, permitindo qualificar entidades espaciais que não têm limites definidos e que são expressos com graus de participação, fazendo com que seja mais apropriada do que uma classificação binária sim/não. A Figura 5.4 exhibe as diferenças entre a lógica booleana e a lógica *fuzzy*.

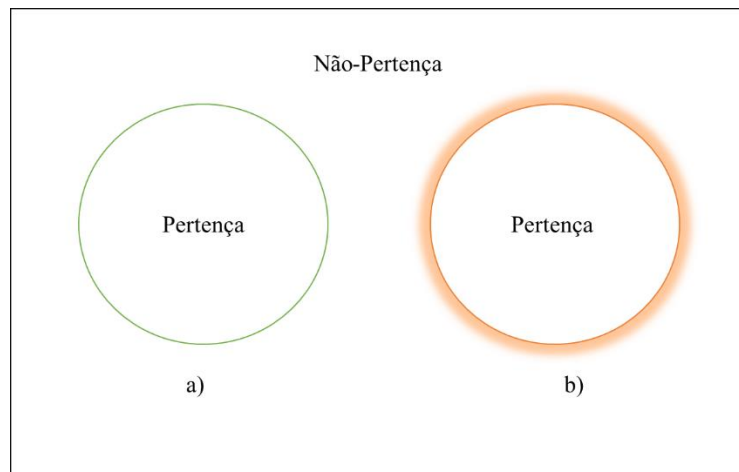


Figura 5.4: Lógica Booleana (a) vs. Lógica Fuzzy (b)

Fonte: Sui (1992, p.103)

Tal como a figura acima mostra, a lógica booleana funciona como uma classificação binária (duas classes mutuamente exclusivas), onde um fenómeno apenas pode ocorrer ou não ocorrer, não sofrendo alterações ao longo do tempo. Por outro lado, a lógica *fuzzy* não tem os seus limites tão rígidos, permitindo que o fenómeno possa ter um escalonamento de valores representando a sua possibilidade de ocorrência, ou seja, a transição entre a pertença e a não-pertença de uma localização num conjunto é gradual.

Por norma, um conjunto *fuzzy* varia entre 0,0 e 1,0, indicando um incremento contínuo da total não-pertença à total pertença. Para representar dados espaciais num esquema *fuzzy*, é necessária a sua transformação para camadas matriciais, uma vez que a maioria dos dados geográficos não possui limites nítidos e cada célula como um conjunto de elementos (Sui, 1992). Assim, a cada célula no ficheiro matricial será anexado um conjunto de classificações de associações que indicam o atributo que estará representado na célula.

Esta abordagem apresenta funções diferentes para definir a transformação ou remapeamento dos valores de entrada para novos valores (Quadro 5.1). O processo de transformação é chamado de *fuzzification*, e estabelece a associação *fuzzy* para cada valor de entrada. Os valores transformados variam de 0.0 a 1.0, definindo a possibilidade de associação a uma classe ou conjunto especificado, em que 1 determina a pertença ao conjunto. Cada classe *fuzzy* define uma função contínua e cada função captura um tipo diferente de transformação para obter um efeito desejado. (ESRI, sem data-a).

Dito isto, para esta dissertação foram transformadas as variáveis, anteriormente tratadas, em classes *fuzzy* através da função *FuzzyLinear*, para assim conseguir realizar os modelos estatísticos preditivos.

Quadro 5.1: Funções Fuzzy (ArcGIS help)

<i>Tipo de Função</i>	<i>Definição</i>
<i>FuzzyGaussian</i>	Define uma função de pertença difusa através de uma distribuição normal ou gaussiana em torno de um ponto médio especificado pelo utilizador (que é atribuído a uma pertença de 1), com um <i>spread</i> definido diminuindo para zero.
<i>FuzzyLarge</i>	Define uma função de pertença difusa, onde os maiores valores de entrada têm pertença mais perto de 1. A função é definida por um ponto médio especificado pelo utilizador (que é atribuído a uma pertença de 0,5), com um <i>spread</i> definido.
<i>FuzzyLinear</i>	Define uma função de pertença difusa através de uma transformação linear entre o valor mínimo especificado pelo utilizador, com uma pertença de 0, até o valor máximo definido pelo utilizador, que é atribuído a uma pertença de 1.
<i>FuzzyMSLarge</i>	Define uma associação difusa através de uma função baseada na média e no desvio-padrão, com os maiores valores tendo uma pertença mais perto de 1.
<i>FuzzyMSSmall</i>	Define uma associação difusa através de uma função baseada na média e no desvio-padrão, com valores menores, tendo uma pertença mais perto de 1.
<i>FuzzyNear</i>	Define uma função de pertença difusa em torno de um valor específico, que é definido por um ponto central indicado pelo utilizador (que é atribuído a uma pertença de 1), com um <i>spread</i> definido diminuindo para zero.
<i>FuzzySmall</i>	Define uma função de pertença difusa com os menores valores de entrada tendo pertença mais perto de 1. A função é definida por um ponto médio especificado pelo utilizador (que é atribuído a uma pertença de 0,5), com um <i>spread</i> definido.

5.3. Modelos estatísticos

A modelação da distribuição dos estabelecimentos comerciais é o processo de construir uma representação dos requisitos fundamentais para uma cadeia de distribuição e extrapolar esses requisitos para uma região geográfica. A importância de poder prever a distribuição das lojas é atualmente destacada pelo aumento do número de cadeias de distribuição existentes e em concorrência direta com a cadeia em estudo.

Os pacotes convencionais de SIG são úteis para a gestão de dados, recolha, visualização e análise espacial, mas não possuem abordagens estatísticas avançadas, particularmente métodos que são relevantes para modelar a distribuição de espécies. No entanto, embora os pacotes estatísticos sejam capazes de analisar e modelar dados de espécies com uma variedade de técnicas de modelagem, a visualização e o suporte a dados SIG são frequentemente deficientes e requerem uma curva de aprendizagem elevada para os utilizadores (Wielanda et al. 2006, citados por Guo & Liu, 2010).

Assim sendo, existem várias abordagens computacionais que podem ser usadas para gerar modelos de distribuição potencial, cada um atingindo resultados ótimos sob diferentes condições. No entanto, os pacotes de *software* existentes disponíveis para esse fim, geralmente, implementam um único algoritmo e cada pacote de *software* apresenta uma nova curva de aprendizagem para o utilizador (Muñoz et al., 2011). Na modelação espacial é comum utilizar-se diferentes técnicas estatísticas para desvendar a complexidade das interações entre distribuições e fatores ambientais. Em 2000, Guisan e Zimmermann, fizeram uma revisão de modelos preditivos e notaram que havia poucos estudos que comparassem mais do que duas técnicas estatísticas aplicadas ao mesmo conjunto de dados.

Atualmente existem muitos pacotes estatísticos que permitem a modelação da distribuição de pontos. No entanto, para esta dissertação optou-se por se utilizar o OpenModeller Desktop. Este *software* foi criado pelo Centro de Referência em Informação Ambiental (CRIA) em 2003, especialmente concebido para estudos ecológicos e ambientais, mas bastante adaptável a diferentes temas, robusto e compatível com várias plataformas. O desktop inclui recursos para amostragem de pontos, criação, teste, avaliação e projeção de modelos em diferentes cenários ambientais, leitura da ocorrência de espécies e dados ambientais em diferentes formatos e recursos (Muñoz et

al., 2011). Deste modo, o OpenModeller apresenta-se como uma ferramenta de modelação de distribuição espacial estática em código aberto (CRIA, sem data).

Este *software* baseia-se na noção de que um Modelo de Distribuição Potencial (Modelo DP) é um modelo matemático gerado por um algoritmo baseado na abordagem correlativa. Esse processo recebe como entrada um conjunto de pontos de ocorrência georreferenciados (presença ou ausência de lojas), um conjunto de dados ambientais (fatores que podem explicar a presença de lojas) georreferenciados e um conjunto de parâmetros específicos do algoritmo. Para criar um Modelo DP, o openModeller lê os valores ambientais correspondentes para cada ponto de ocorrência a partir dos ficheiros matriciais de entrada. Os pontos de ocorrência são transformados em estruturas de dados, chamadas de amostras, cujos elementos representam as condições ambientais em cada local. O conjunto de todas as amostras é então usado por um algoritmo para encontrar uma representação das lojas no espaço ambiental. O resultado é um modelo DP, que pode ser um modelo de dados probabilístico ou uma função matemática que relaciona as condições ambientais à adequação para a existência de espécies (Muñoz et al., 2011).

Aqui é possível encontrar dois tipos de modelos: os mais fáceis de interpretar, como os algoritmos *envelope-based*, como o Bioclim, e os algoritmos de *Machine Learning* como as *Decision Trees*. No entanto, a maioria dos modelos são de difícil interpretação, como é o caso do *Genetic Algorithm for Rule-set Prediction* (GARP), as *Support Vector Machines* (SVM) e as *Artificial Neural Networks* (ANN). Geralmente os modelos DP não são interpretados, mas usados para projetar o modelo numa região geográfica de destino. Nesse processo, as condições ambientais são lidas iterativamente a partir de um conjunto de ficheiros matriciais para cada posição da célula associada a uma região de destino. Este conjunto de células pode ser o mesmo usado para gerar o modelo DP (projeções nativas), sendo também possível utilizar ficheiros matriciais relacionados a um período diferente no tempo ou numa região geográfica.

O conjunto de condições ambientais para cada célula é então passado como um parâmetro para o algoritmo. O algoritmo retorna um valor de previsão correspondente à adequação deste ambiente para a abertura de novas lojas. Cada previsão é então gravada numa célula correspondente no mapa de saída, georreferenciada, e o resultado é um Mapa de Distribuição de Potencial (mapa DP). Um mapa DP representa a distribuição potencial das lojas numa região geográfica específica num determinado momento. Os valores de previsão podem ser probabilidades reais ou podem ser o resultado de uma função que

devolve valores contínuos ou categorizados, dependendo do algoritmo usado. O significado dos valores de previsão num mapa DP só pode ser interpretado corretamente pelo algoritmo que produziu o modelo DP (Muñoz et al., 2011).

Tendo as últimas informações em consideração, foi necessário preparar-se dois conjuntos de dados. Em primeiro lugar, os ficheiros matriciais, escalonados de 0 a 1 (lógica *fuzzy*), anteriormente criados nas variáveis a entrar nos modelos, foram convertidos para o sistema de coordenadas WGS84 (sistema de coordenadas aceite pelo openModeller) e colocados numa pasta para que fosse possível a sua utilização.

Em seguida, preparou-se o ficheiro de ocorrência (formato texto) com as lojas Meu Super, as suas coordenadas e informação sobre a ausência ou presença das lojas, neste caso, presença (Figura 5.5).

#ID	Taxon Name	Longitude	Latitude	Abundance
1	152410	-9.379619	38.803438	1
1	152426	-8.428305	40.522028	1
1	152409	-9.296438	38.774342	1
1	152568	-9.352602	38.758451	1
1	152577	-8.156495	38.393663	1
1	152400	-8.769254	39.468812	1
1	152565	-9.004245	39.051237	1
1	152550	-8.614020	41.178647	1
1	152412	-9.229500	38.704449	1

Figura 5.5: Exemplo do ficheiro de ocorrências (nota: as casas decimais estão representadas por pontos devido às especificidades do *software*)

Com os dados preparados, o passo seguinte passou pela escolha dos algoritmos a usar para a criação dos mapas. Essa escolha passou pela opção de demonstrar diferentes graus de “dificuldade” de interpretação dos modelos existentes. Os algoritmos escolhidos nesta dissertação foram: *Environmental Distance* (com a distância euclidiana) baseado na distância; GARP, baseado nos algoritmos genéticos; e as SVM, baseadas em redes neurais.

5.3.1. *Environmental Distance*

O algoritmo *Environmental Distance* veio substituir os anteriores algoritmos baseados em distâncias no openModeller (*Minimum Distance* e *Distance to Average*),

apresentando uma abordagem mais geral. Este algoritmo é uma coleção totalmente determinista de algoritmos de previsão baseados na distância, no entanto, apesar de este ser um algoritmo determinístico os seus resultados são valores de probabilidade.

É necessário imaginar um mundo bidimensional onde cada eixo é uma possibilidade ambiental, o que significa que um ponto de amostragem será visto como um valor num espaço geométrico dimensional parecido com o Euclidiano. No espaço usado para uma experiência é possível calcular uma distância entre dois pontos usando uma equação (normalmente, a norma do vetor de diferença). Existem várias possibilidades de equações, a sua escolha vai estar dependente da utilidade que se dá ao modelo.

Todas as distâncias são calculadas entre dois pontos e é necessário encontrar uma probabilidade para cada ponto dado pelo openModeller ao algoritmo. Isso significa que é necessário encontrar um ponto que represente o lugar onde a distância deve ser calculada para cada ponto que o *software* fornece. Para isso, o algoritmo usa os pontos de presença dados pelo utilizador.

Assim, este algoritmo apresenta-se como um algoritmo genérico baseado em métricas de disparidades ambiental, que utiliza apenas os pontos de presença fornecidos pelo utilizador, ignorando os pontos de ausência e sem necessidade de criar pontos de “pseudo-ausências”, ou seja, não requerendo conhecimento prévio da origem e capacidade de dispersão da espécie para a calibração do modelo. O algoritmo permite aos utilizadores especificar uma distância métrica a usar (Euclidiana, Gower, Mahalanobis ou Chebyshev), o número de pontos mais próximos a serem tomados como referência e a distância máxima até o centroide dos pontos mais próximos.

Para cada ponto, o algoritmo encontra um novo ponto, correspondente ao ponto médio dos “n” pontos mais próximos de um dado ponto, para calcular uma distância e utiliza essa distância como base para calcular um valor de probabilidade.

A Figura 5.6, ilustra a explicação do algoritmo. Os pontos alaranjados representam os pontos de presença num mundo bidimensional, e o ponto mais escuro é o ponto dado pelo openModeller (oM). Deste modo, o algoritmo calcula as distâncias entre o ponto dado e cada ponto de presença, e, assim, coloca todos os pontos de presença em ordem, do ponto mais próximo para um mais distante. Por último, o algoritmo utiliza o parâmetro que dá o valor “n” dos pontos mais próximos que devem ser usados. Por exemplo, se “n”

for 3, o algoritmo encontra a média dos 3 pontos mais próximos e usa a distância entre este e o ponto dado. Na Figura 5.6 é apresentada a média dos pontos alaranjados (a e b) e a distância que é usada para obter o valor da probabilidade no *output* (c).

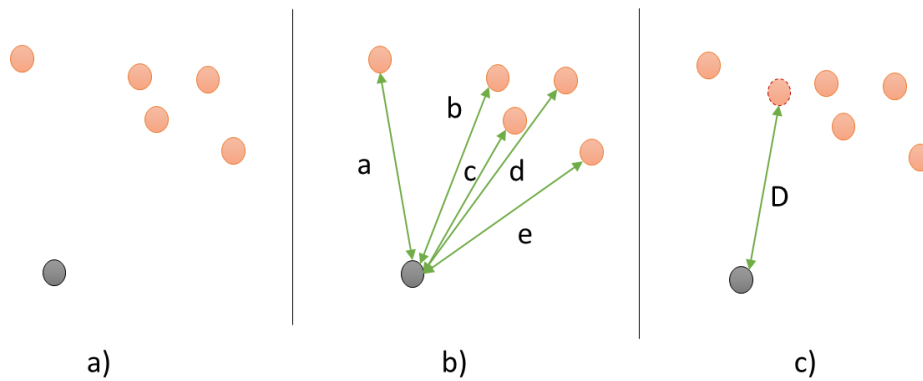


Figura 5.6: Algoritmo Environmental Distance

Fonte: Canhos (2009, p.42-43)

O valor de “n” terá de ser um valor inteiro entre $[1; p]$, em que p corresponde à quantidade de pontos de presença usados no modelo. O intervalo de distância começa em zero, onde o ponto dado pelo *software* é exatamente o ponto médio calculado, mas o máximo deve ser calculado usando um algoritmo. Para ajudar os cálculos, o intervalo de cada dimensão pode ser normalizado para ser sempre $[k1; k2]$, portanto, todos os possíveis pontos estão dentro de um hipercubo. Para a distância euclidiana, a distância máxima é a distância entre dois vértices opostos do hipercubo (quadrado, neste caso). A probabilidade é calculada linearmente, usando um intervalo de distância. Supondo que o utilizador não tenha colocado um limite de saturação para a distância máxima, a faixa de distância usada pela avaliação de probabilidade é exatamente a faixa de distância observada na Figura 5.6c) (Canhos, 2005).

Na verdade, a linha é uma curva do tipo escada, já que o oM é um *software* que trabalha com dados digitais e que, de alguma forma, os dados precisam ser quantificados e podem ser uma fonte de erros. Em alguns casos, a distância pode estar sempre num pequeno intervalo próximo da probabilidade de 1, ou mesmo de num intervalo que não altera a probabilidade, originando modelos sem informação. Para isso, o que pode ser feito é o que está representado na Figura 5.7.

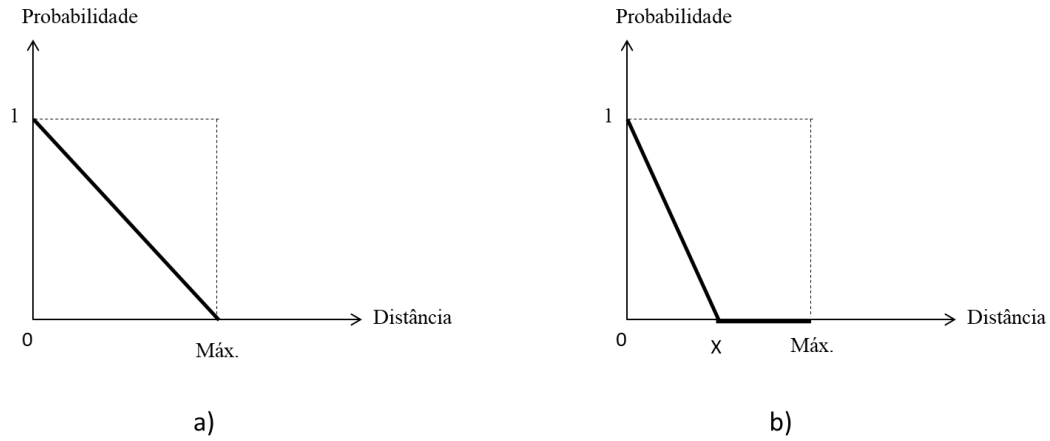


Figura 5.7: Probabilidade do modelo

Fonte: Canhos (2009, p.43-44)

A probabilidade satura em zero, na Figura 5.7b, uma vez que se usa X como distância máxima, e não o máximo. Com isso, o intervalo de probabilidade pode ser controlado para dar valores com representação significativa, assim o utilizador não precisa de se preocupar com o erro de quantização do modelo como antes. O valor de X é sempre uma fração do valor máximo, sendo essa fração definida pelo utilizador como um parâmetro (com a distância máxima de um), a que se chamará de "m".

Dito isto, utilizando D como a distância descoberta, a equação da curva acima [0;X] é dada por:

$$Probabilidade = 1 - \frac{D}{X} \quad (5.3)$$

Ou

$$Probabilidade = 1 - \frac{D}{m * Max} \quad (5.4)$$

Existem três parâmetros usados pelo algoritmo: um que informa a métrica que deve ser utilizada (euclidiana, por exemplo), o outro é a quantidade de pontos que são usados para encontrar um ponto médio, e o último é um ponto saturação máxima de distância, para evitar modelos nulos como *output*.

5.3.1.1. Distância euclidiana

Cada distância ilustrada antes foi calculada usando a mesma equação, escolhida pelo utilizador como um parâmetro. Para as distâncias não importa como o algoritmo calcula as probabilidades, só precisa dos dois pontos que serão usados. Assim, as equações de distância têm dois pontos de amostra como entrada e um valor contínuo como saída.

A distância euclidiana, conhecida também como a segunda norma da distância Minkowski ou a Generalização do Teorema de Pitágoras, é a distância mais comum utilizada na geometria, na física, na matemática, etc. Esta pode ser calculada utilizando a norma do vetor da distância, sendo facilmente calculada com a subtração de um ponto da amostra a outro ponto. A norma é a raiz quadrada da soma dos quadrados da coordenada de cada vetor, ou simplesmente a raiz quadrada do produto de ponto do vetor com ele mesmo (Canhos, 2005).

Dito isto, procedeu-se à aplicação do modelo, que originou o resultado da Figura 5.8. Através desta figura, é possível observar a flexibilidade deste algoritmo. Este criou áreas, envolventes das lojas existentes, de aproximadamente 15km, diminuindo a aptidão à medida que se afasta das lojas. Devido a esta especificidade do algoritmo, mais de 60% do território apresenta aptidão muito reduzida (0-0,3) para a localização dos novos estabelecimentos, e apenas 7% apresenta aptidão elevada (aptidão superior a 0,7). Nota-se uma grande concentração de locais com aptidão elevada nas áreas metropolitanas de Lisboa e Porto, assistindo-se ao inverso onde não existem lojas Meu Super, uma vez que quanto mais longe dos pontos de amostra (lojas MS) menor a aptidão para a localização de novos estabelecimentos.

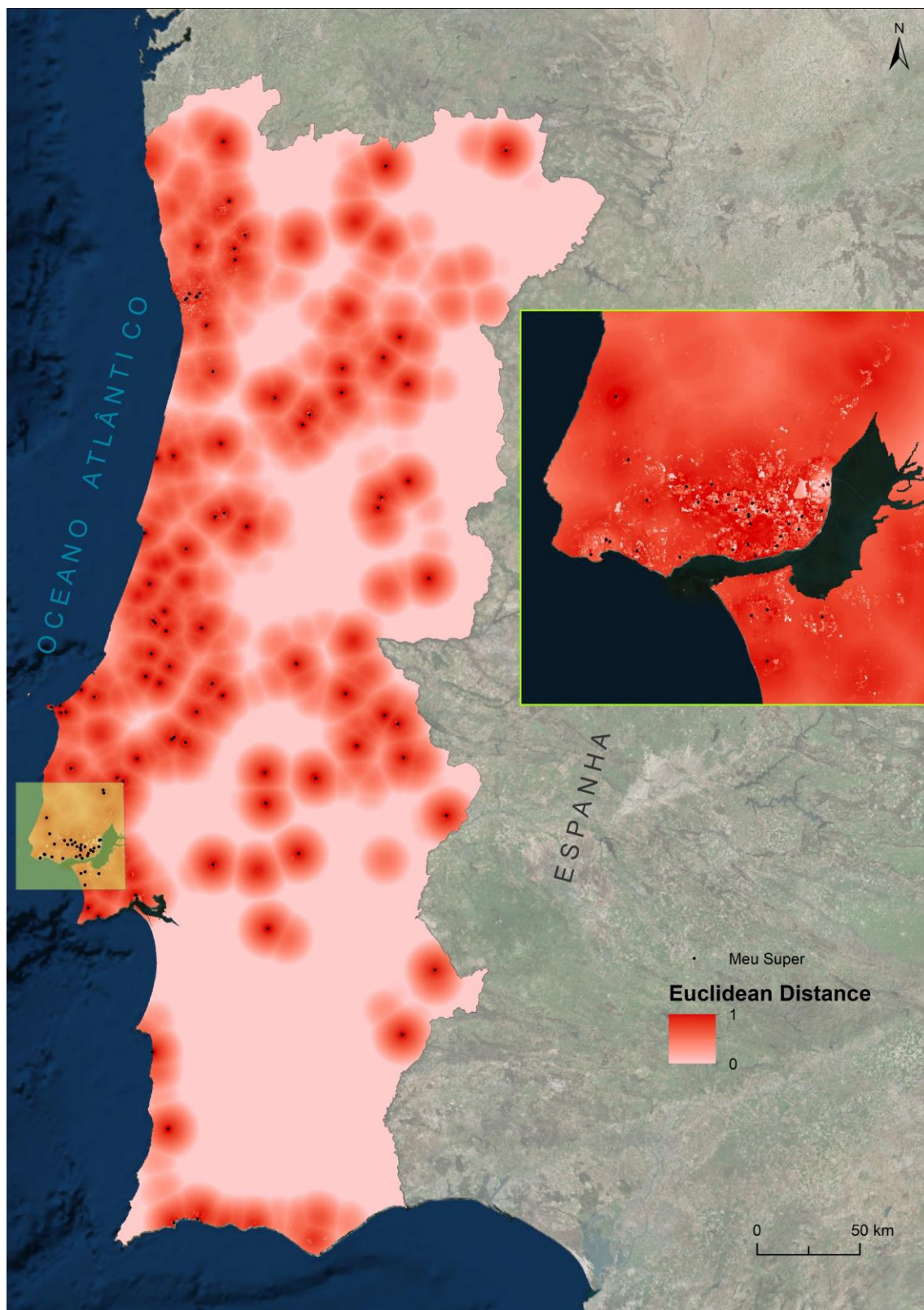


Figura 5.8: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a aplicação algoritmo *Environmental Distance* (através da distância euclidiana)

5.3.2. *Support Vector Machines*

Nos últimos anos, com o avanço da eficiência computacional combinada com a sofisticação métodos estatísticos, os métodos *Machine Learning* têm sido, cada vez mais, utilizados como ferramentas poderosas numa ampla variedade de disciplinas científicas, incluindo a ciência aeronáutica, a ciência da computação, a bioinformática e a ciência ambiental (Mjolsness e DeCoste, 2001, citados por Guo et al., 2005). Entre os métodos *Machine Learning*, as máquinas de suporte vetorial (*Supported Vector Machines* –SVM), desenvolvidas originalmente por Vapnik (1995), são consideradas uma nova geração de algoritmos de aprendizagem, apresentando características atraentes para os modeladores, incluindo a sua base em estatística, em vez de analogias fracas com sistemas de aprendizagem naturais, e apresentam, teoricamente, melhor desempenho (Guo et al., 2005).

Este algoritmo caracteriza-se por conter um conjunto de métodos de aprendizagem supervisionados relacionados que pertencem à família dos classificadores lineares generalizados (Vapnik, 2005, citado por Barreto, 2008), apresentando a capacidade de introdução na ordem de 100 000 amostras, criando hiperplanos de separação que atuam como classificadores dos dados, gerando probabilidades de presença e ausência.

Por norma, as SVM são projetadas para problemas de duas classes em que existem objetos positivos e negativos. Para esses problemas de classificação, a SVM de duas classes procura localizar um hiperplano no espaço que separe as duas classes de destino. No entanto, muitas vezes não temos dados negativos e, portanto, existem apenas dados para um conjunto de dados de uma classe. Computacionalmente, este é um problema difícil de resolver, uma vez que com duas classes é possível “desenhar” o hiperplano, conseguindo treinar e restringir os modelos estatísticos de aprendizagem, o mesmo não acontece com apenas uma classe.

De modo a mapear os vetores de entrada (geralmente pontos) num espaço bidimensional, as SVM constroem um hiperplano de separação máxima, de modo a separar os dados no espaço. Em seguida são construídos mais dois hiperplanos paralelos ao hiperplano de separação, sendo este o que maximiza a distância entre os dois hiperplanos paralelos (Figura 5.9). Pressupõe-se que quanto maior a margem, ou a distância, entre esses hiperplanos paralelos, menor será o erro de generalização do classificador. O modelo produzido pela classificação de vetor de suporte depende apenas

de um subconjunto dos dados de treino, porque a função de custo para construir o modelo não se preocupa com os pontos da amostra de treino que estão para além da margem.

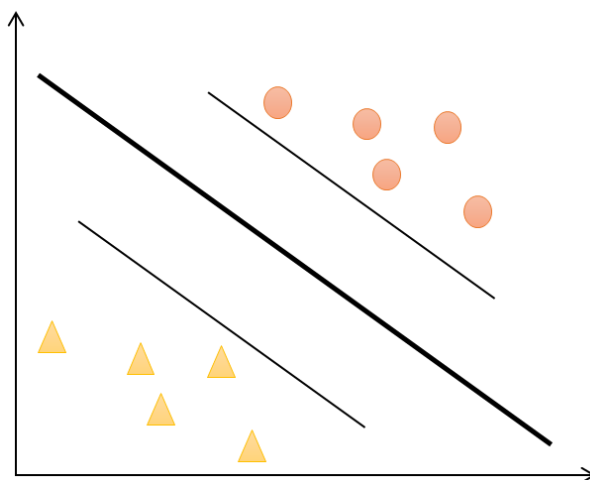


Figura 5.9: Hiperplanos das SVM

As SVM são técnicas de *Machine learning* baseadas em conceitos estatísticos, utilizados em diferentes aplicações que envolvem a descoberta de padrões em dados espaciais. Estas apresentam-se como uma classe de algoritmos de reconhecimento de padrões estatísticos não-probabilísticos para estimar, entre outras coisas, o limite do conjunto a partir do qual um grupo de observações é criada. Os métodos *Support Vector Machines* (SVM) identificam, assim, um envelope ambiental ou hiperespaço contendo os pontos de dados, de acordo com os quais o envelope é otimizado em função do número que se encontra dentro e fora (*outliers*) do envelope. Dito isto, aplicou-se este algoritmo aos dados relativos ao Meu Super resultando a Figura 5.10.

Com a Figura 5.10, é possível observar a reduzida flexibilidade deste algoritmo. Este criou áreas, envolventes das lojas existentes, de aproximadamente 3km, diminuindo a aptidão à medida que se afasta das lojas. Devido a esta especificidade do algoritmo mais de 96% do território apresenta aptidão muito reduzida (0-0,3) para a localização dos novos estabelecimentos, e apenas 1,7% apresenta aptidão elevada (aptidão superior a 0,7). É possível observar a influência de algumas das variáveis utilizadas no modelo (*e.g.*: densidade de estradas ou densidade de lojas da concorrência), com especial foco nas áreas metropolitanas de Lisboa e Porto, onde existiam muitos fatores que influenciavam a aptidão para a localização dos novos estabelecimentos, e onde as variáveis apresentam correlação forte entre elas.

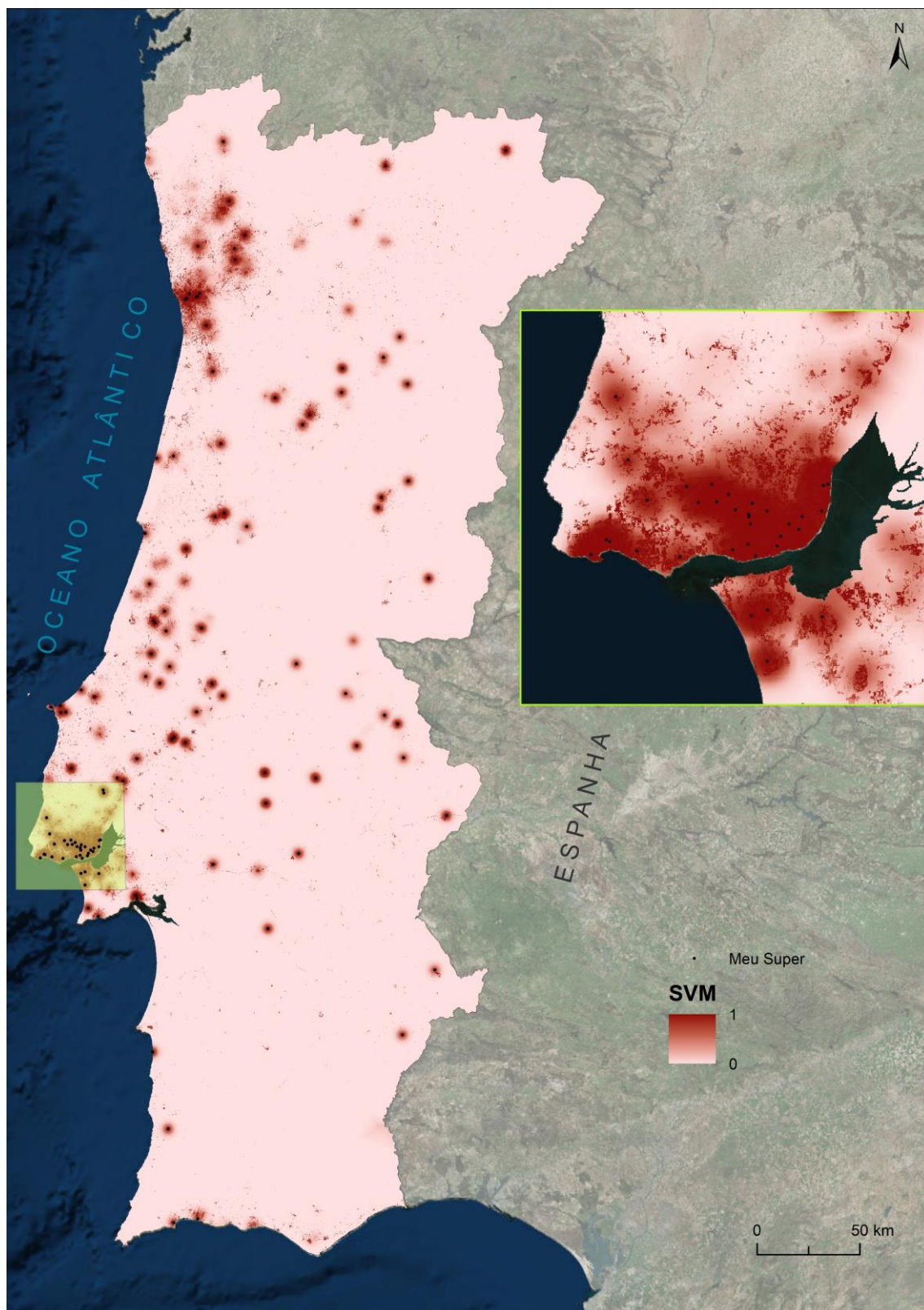


Figura 5.10: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a aplicação do algoritmo *Support Vector Machines*

5.3.3. *Genetic Algorithm for rule-set prediction (GARP)*

O GARP é, à semelhança das SVM, um método de *Machine Learning*. Este utiliza um algoritmo genético para selecionar um conjunto de regras (*e.g.*, regressão logística, regras bioclimáticas, etc.) que melhor predizem a distribuição das variáveis dependentes (Stockwell & Peters, 1999). Sendo um algoritmo genético, este é inspirado no conceito de evolução através da seleção natural e baseia-se na ideia de desenvolver soluções para os problemas de uma forma análoga à forma como os organismos evoluem. Fazendo com que o objetivo principal seja criar um conjunto de possíveis soluções para um problema (a distribuição das lojas) e depois modificar iterativamente e testar esse conjunto até que uma solução ótima seja encontrada.

Os modelos descrevem as condições ambientais sob as quais poderão ser abertas novas lojas MS. Para a entrada, o GARP usa um conjunto de localizações pontuais onde já existem lojas e um conjunto de camadas geográficas representando os parâmetros ambientais que podem limitar as capacidades para a abertura de novas lojas. Esse algoritmo aplica o procedimento de melhores subconjuntos usando a nova implementação de openModeller em cada execução de GARP.

Os modelos GARP foram desenvolvidos usando uma versão desktop que permite flexibilidade no desenvolvimento de modelos (Pereira, 2002; citado por Townsend Peterson et al., 2007). No GARP, os pontos de ocorrência do par de quadrantes nos quais os modelos serão baseados são divididos igualmente em conjuntos de treino e de "dados de teste extrínsecos"; sendo o conjunto anterior, novamente dividido igualmente em verdadeiros conjuntos de "dados de treino" (para desenvolvimento de regras de modelos) e "dados de testes intrínsecos" (para avaliação e refinamento de regras de modelos) (Townsend Peterson et al., 2007).

Este modelo funciona num processo iterativo de seleção de regras, testes e incorporação ou rejeição. O GARP começa por escolher um método de um conjunto de possibilidades (regressão logística, regras bioclimáticas, regras de intervalo, regras de intervalo negadas), sendo aplicado nos dados de treino e desenvolvida uma regra; as regras podem evoluir de várias maneiras (por exemplo, truncamento, alterações pontuais, *crossing-over* entre regras) para maximizar a previsão.

Dos modelos GARP disponíveis no openModeller, foi aplicado o procedimento de "melhor subconjunto" (*best subsect*) que seleciona os melhores modelos com base nas

estatísticas de omissão e erro de comissão. Este foi originalmente desenvolvido para filtrar a variação de modelo-a-modelo gerada pela natureza aleatória do algoritmo GARP. Com este modelo foram gerados modelos para cada espécie e foram selecionados os "melhores" baseados em omissões (excluindo áreas de potencial distribuição) e comissões de estatísticas de erros (incluindo áreas potencialmente não habitáveis) (Peterson et al. 2007) (Townsend Peterson et al., 2007).

Exposto isto, procedeu-se à aplicação deste modelo aos dados disponíveis, obtendo-se o resultado da Figura 5.11.

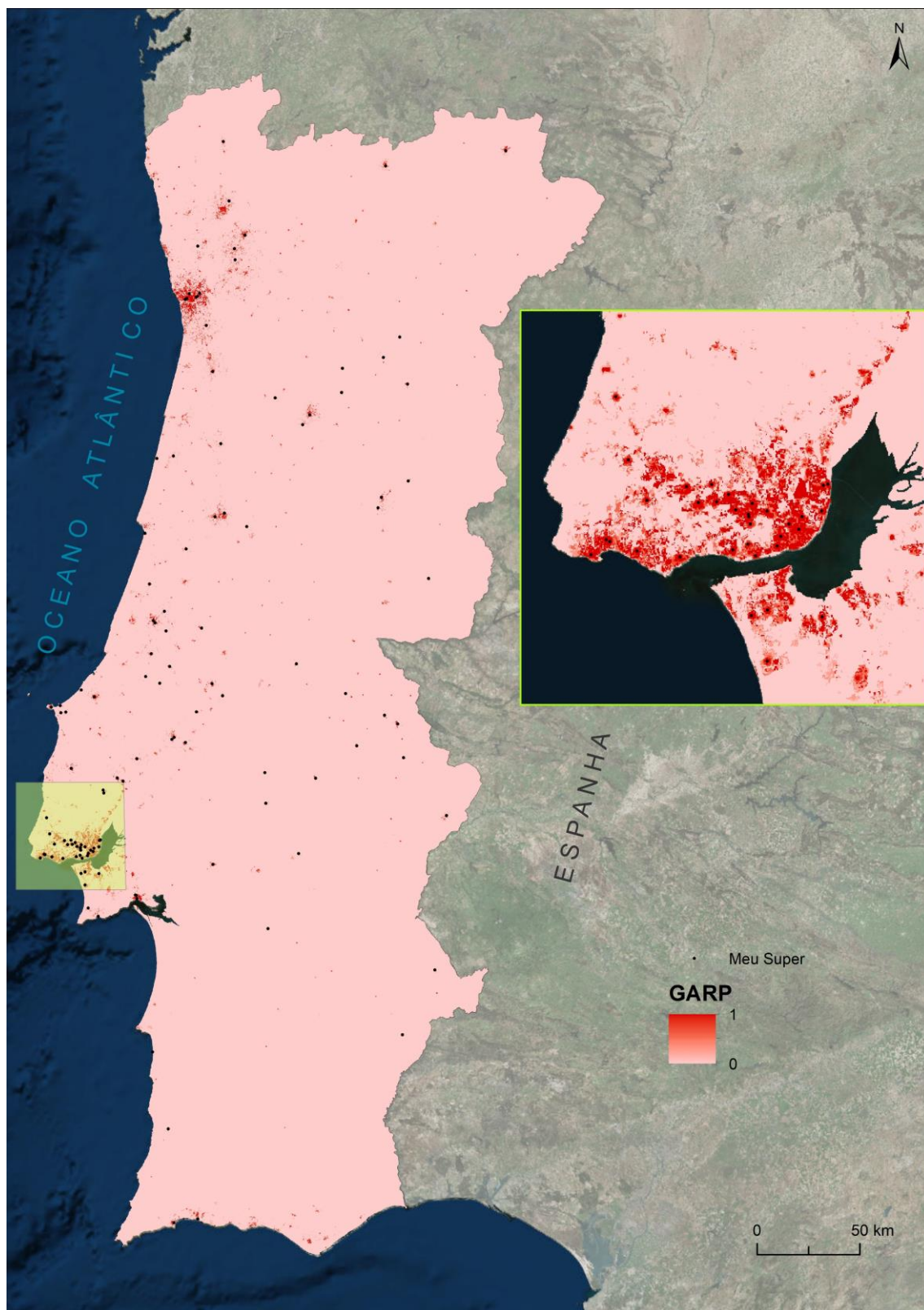


Figura 5.11: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a aplicação do algoritmo GARP

Com a análise da Figura 5.11, é possível notar a falta de flexibilidade deste algoritmo. Este criou áreas, envolventes das lojas existentes, de aproximadamente 500m, diminuindo rapidamente a aptidão à medida que se afasta das lojas. Devido a esta especificidade do algoritmo mais de 99% do território apresenta aptidão muito reduzida (0-0,3) para a localização dos novos estabelecimentos, e apenas 0,2% apresenta aptidão elevada (aptidão superior a 0,7). É possível concluir, que tal como os restantes modelos, este também apresenta maior concentração de locais com aptidão elevada nas áreas metropolitanas de Lisboa e Porto, e que as áreas determinadas com maior aptidão correspondem, na sua maioria, a áreas onde já existem lojas MS, assim como a presença de lojas Pingo Doce.

5.3.4. Aptidão

Foram apresentados três resultados distintos de aptidão à localização de novos estabelecimentos comerciais de proximidade, no entanto não há um resultado melhor que outro, visto que é uma análise preditiva e todos apresentam incertezas. Assim sendo, optou-se por fazer a média dos três resultados para obter um mapa que permite diminuir a variabilidade entre os resultados, tornando-o mais flexível e generalizado (Figura 5.12).

Através da Figura 5.12, consegue-se observar a influência dos mapas anteriores, neste novo resultado. Este criou áreas de aptidão de 15km, provenientes da Distância Euclidiana, no entanto apenas as áreas de 500 m do GARP apresentam maior aptidão à localização de novos estabelecimentos, e é possível observar ainda a influência do SVM nas pequenas áreas de aptidão, dispersas pelo território. Devido a esta agregação dos algoritmos, mais de 95% do território apresenta aptidão muito reduzida (0-0,3) para a localização dos novos estabelecimentos, e apenas 0,3% apresenta aptidão elevada (aptidão superior a 0,7).

É de salientar ainda, que apesar de haver variáveis que têm valores inversos, ou seja, variáveis que apresentavam valores negativos onde outras apresentavam valores positivos, como é o caso da variável “densidade de lojas MS”, onde os locais perto das lojas apresentam aptidão reduzida, nos resultados de aptidão estas áreas apresentam aptidão elevada. Isto deve-se ao facto de os modelos funcionarem com as correlações entre os dados, ou seja, se os dados apresentarem forte correlação nesses locais (quer correlação positiva quer negativa) vão apresentar aptidão elevada.

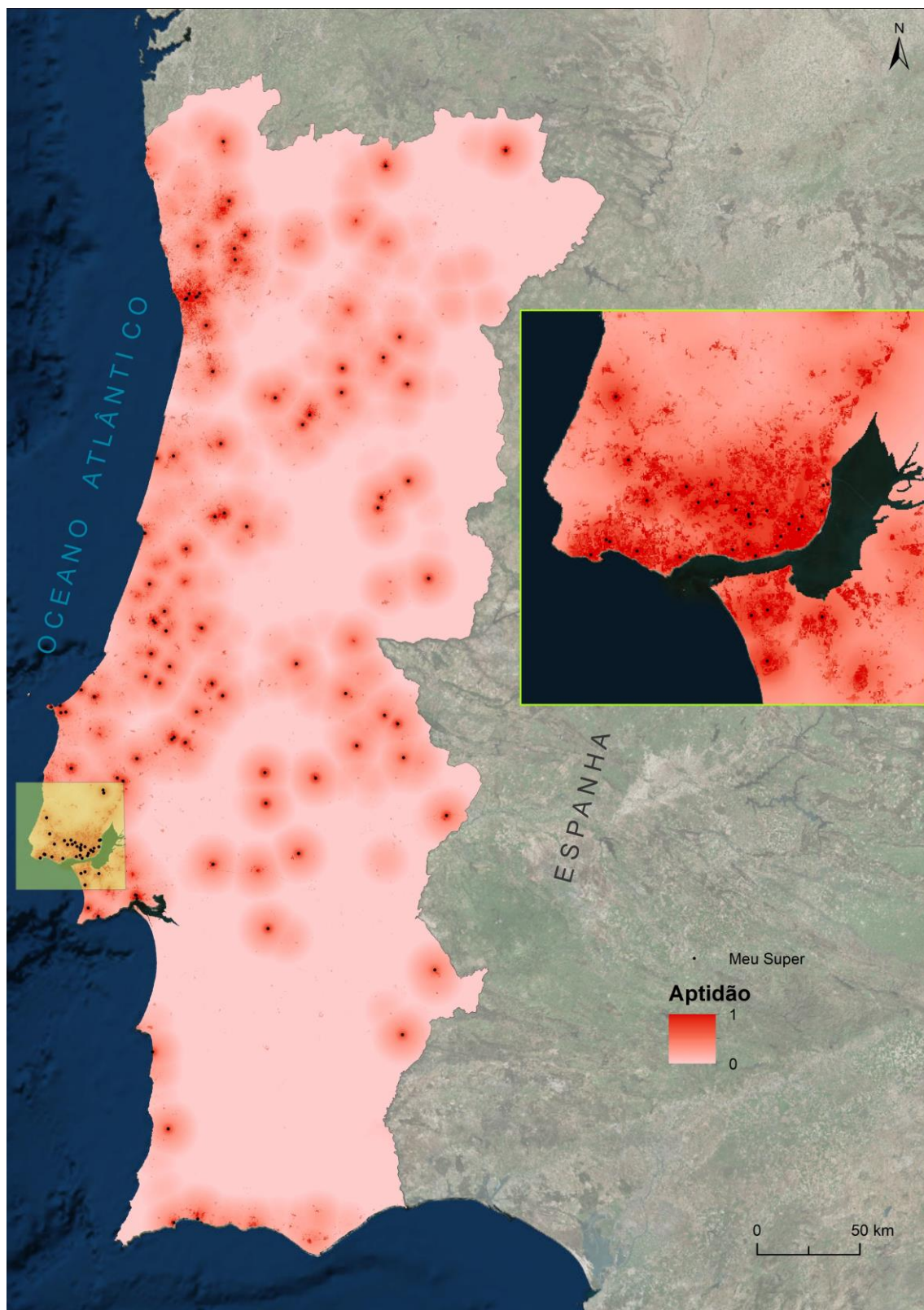


Figura 5.12: Aptidão espacial para a implantação de novas lojas Meu Super, de acordo com a média dos algoritmos aplicados

5.3.1. Validação

A qualidade dos dados SIG é importante. Quando criamos ou atualizamos dados, fazemo-lo manualmente, o que pode levar à ocorrência de erros. Esses erros podem originar resultados sem qualidade e pouco realistas, o que poderá ter ramificações extremas. Antes de se tentar usar um conjunto de dados para alguma tarefa importante, todos os aspetos devem ser verificados quanto à integridade, correção, consistência e conformidade. Para se ter dados de boa qualidade é necessário verificar se estes correspondem aos requisitos e, em seguida, repará-los onde eles não se adequam.

Assim sendo, com a aptidão definida, tornou-se importante validar os resultados, recorrendo-se, para isso, às lojas Meu Super encerradas para determinar se as lojas que encerraram estão em áreas a que foi atribuída menor aptidão para a localização de novos estabelecimentos.

Aquando da extração dos valores de aptidão das lojas encerradas (Quadro 5.2), deparamo-nos com dois cenários. O primeiro em que sete das lojas encerradas apresentavam baixa ou média aptidão, o que se esperava, podendo ser explicado por estarem localizadas em áreas remotas e com pouca densidade populacional. E o segundo, em que duas lojas apresentam aptidão elevada, mas que, no entanto, foram encerradas, podendo ser explicado por diversos fatores, entre eles, o perfil do comerciante, o perfil do consumidor ou a existência de concorrentes na mesma área. Como não existe informação auxiliar referente aos comerciantes ou consumidores nessas áreas, teve-se especial atenção à distribuição espacial dessas lojas.

Quadro 5.2: Aptidão das lojas encerradas - validação

Nome da Loja	Aptidão
Meu Super Coimbra, I	0,86
Meu Super Porto	0,74
Meu Super Vermoil (Pombal)	0,60
Meu Super Porto - Constituição	0,58
Meu Super Águeda	0,57
Meu Super Massamá	0,51
Meu Super Arazede	0,47
Meu Super Ovar	0,31
Meu Super Beja	0,00

Em Coimbra, a loja encerrada (ponto vermelho) encontrava-se num raio de 2km de outras duas lojas (pontos amarelos), numa área de densidade populacional mais baixa do que nas lojas já existentes, o que pode explicar o encerramento desta. Em relação às lojas encerradas na cidade do Porto, estas já se encontravam num raio de 2,5 km de outras lojas MS. Esta proximidade a outras lojas da mesma insígnia poderá levar à canibalização da marca, o que é desaconselhável (Figura 5.13).

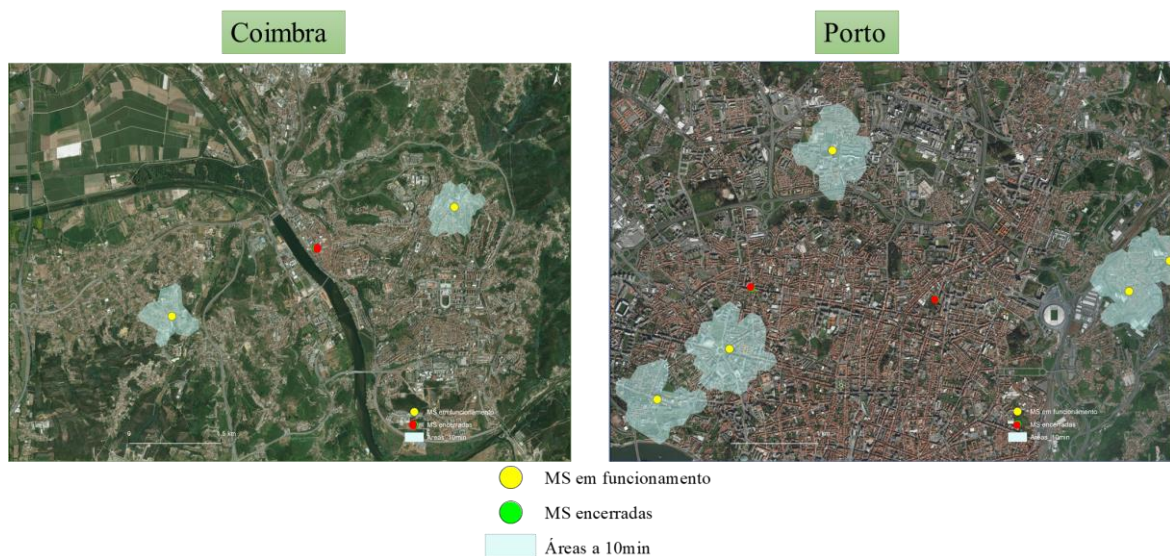


Figura 5.13: Lojas encerradas com aptidão elevada

5.4. Modelos de Localização-Alocação

O planeamento de uma localização, *i.e.*, a determinação do número e da localização de determinados serviços numa rede, é de natureza estratégica e de alta responsabilidade. A decisão de localizar determinado serviço é um processo muito complexo e envolve a alocação de quantias consideráveis de recursos financeiros.

A seleção de um local conveniente e lucrativo para o estabelecimento de uma atividade comercial é um processo complexo e arriscado, que exige uma enorme base de dados dos consumidores para que a análise da localização seja exata. De acordo com a literatura sobre o tema, para encontrar as localizações ótimas, o primeiro pressuposto é

que as lojas devem estar o mais próximas possível dos consumidores, o que confere ao problema um elevado grau de complexidade.

Os SIG, e a possibilidade de integrar dados espaciais e alfanuméricos, vieram permitir a utilização de novos métodos de processamento e forneceram uma alta qualidade na visualização dos dados processados. Estas características tornam os SIG numa ferramenta essencial para a tomada de decisão, uma vez que os problemas de localização tornam necessária a introdução da componente espacial na decisão. Apesar de todos os benefícios, os SIG não proporcionam uma solução final, apenas fornece possibilidades para uma análise mais organizada da informação utilizada.

Grande parte dos modelos cujo objetivo é a procura de locais comerciais ideais têm por base a análise da repartição dos clientes potenciais no espaço, como é o caso dos modelos gravitacionais, *i.e.*, a lei de Reilly e o modelo MCI (modelo de competição interativa), que procuram localizar os estabelecimentos o mais próximo possível dos clientes, de modo a atrair o maior número de clientes. Por outro lado, os modelos de localização-alocação procuram encontrar os locais mais próximos dos clientes através de algoritmos matemáticos (Baray & Cliquet, 2003).

Os modelos localização-alocação foram criados como resposta a um duplo objetivo da teoria das localizações: encontrar localizações ótimas e determinar a alocação da procura por determinado bem ou serviço (Buzai, 2013), combinando assim os fatores essenciais para qualquer bem e serviço, nomeadamente, a procura e a oferta. O objetivo consiste na identificação das melhores localizações e avaliação do conjunto de localizações utilizando uma determinada função (Comber et al., 2015; Gerrard & Church, 1996). Dado um conjunto de instalações que irão abastecer um determinado conjunto de pontos de procura, os modelos localização-alocação vão encontrar os melhores locais tendo em conta fatores como o número de instalações disponíveis, o seu custo e a impedância máxima de uma instalação para um ponto de procura.

De acordo com Ramirez e Bosque (2001, p.57-58), os modelos localização-alocação são:

- Modelos matemáticos, uma vez que esta linguagem é considerada apropriada para capturar a realidade;
- Modelos espaciais de escala intermédia, pois os aspetos a serem resolvidos já estão delimitados num território;

- Modelos normativos, porque é necessário procurar a melhor solução para um dado problema.

De acordo com Goodchild (1984), o sector retalhista apresenta três questões gerais que devem ser abordadas antes da escolha do modelo mais indicado. Primeiramente, a alocação é um problema para o retalho devido à sua grande complexidade e falta de modelos precisos. Em seguida, as funções-objetivo devem refletir as preocupações do sector para a rentabilidade e competitividade. E finalmente, qualquer estratégia de localização deverá ter em conta questões como a rapidez, o custo da análise e os aspetos práticos. Nesta dissertação, por falta de dados mais precisos, apenas serão analisadas as primeiras duas questões.

No setor da venda retalho, apenas as localizações podem ser controladas, uma vez que a alocação da procura é o resultado do comportamento espacial dos clientes, e visto que estes são livres de escolher que lojas visitam, não existe uma teoria exata. Muitas vezes os consumidores dirigem-se às lojas mais próximas de si para satisfazerem as suas necessidades imediatas, optando algumas vezes por uma marca em concreto. No entanto, este padrão é afetado pelas influências de outros tipos de compras como, *e.g.*, as compras impulsivas.

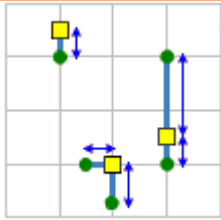
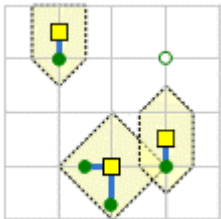
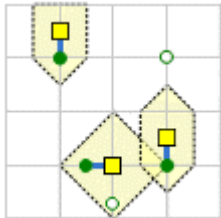
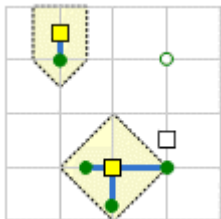
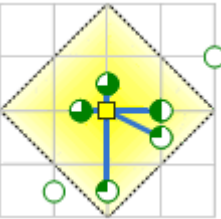
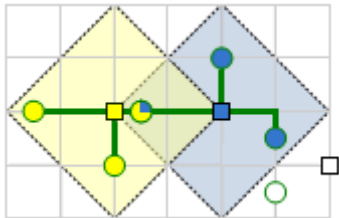
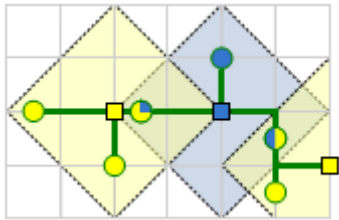
Devido a estas características e ao facto que para cada pessoa cada loja tem diferentes níveis de atratividade, os modelos que atentam modelar a alocação dos consumidores não são exatos. Os modelos localização-alocação não fogem à regra. Para esta dissertação foram utilizados os centroides das áreas urbanas em Portugal continental, como pontos de procura dos estabelecimentos comerciais, e o pressuposto de que os consumidores se irão deslocar para a loja mais próxima de si.

Este modelo tem como único objetivo a eficiência espacial (Ramirez & Bosque, 2001), sendo para isso utilizadas funções-objetivo que confirmam os objetivos a serem otimizados. No Quadro 5.3 estão presentes as funções-objetivo que se podem utilizar quando se fazem modelos localização-alocação.

Quando se aplica estes modelos ao sector do retalho existem duas opções: (1) ignoram-se os concorrentes, optando apenas pela utilização das distâncias entre determinados estabelecimentos e a procura por eles; ou (2) utilizam-se a concorrência de modo a que o modelo fique mais realista e que abarque todos os fatores que influenciam a procura.

A primeira opção representa um método mais básico, no entanto mais agressivo, uma vez que apenas tem em consideração as lojas em estudo e não tem em conta as preferências dos consumidores, minimizando as distâncias entre a procura e as lojas do cliente, selecionando os locais mais acessíveis aos consumidores. A segunda opção já tem em conta os concorrentes das lojas em estudo e é mais apropriada quando os consumidores preferem o local mais próximo, independentemente da marca disponível, podendo até utilizar as lojas de conveniência ou as estações de gasolina para fazer as suas compras. Nesse caso, uma função objetiva apropriada seria maximizar a procura total alocada às lojas do cliente ou a quota de mercado. No entanto, esta função é limitada àquela fração de viagens de compras de conveniência que se originam de pontos fixos, como residências ou locais de trabalho. Assume-se que os locais de lojas competitivas são conhecidos e que a função objetivo deve ser otimizada selecionando localizações apenas para a insígnia em estudo.

Quadro 5.3: Funções- objetivo dos modelos localização-alocação (ArcGIS help)

<i>Ilustração</i>	Designação	Descrição
	Minimização da impedância (p-mediana)	Este método escolhe instalações de modo a que a soma das impedâncias ponderadas (procura atribuída a uma instalação multiplicada pela impedância à instalação) seja minimizada
	Máxima cobertura	O método de máxima cobertura escolhe instalações de modo a que seja abrangida tanta procura quanto possível, delimitada pelo corte de impedância
	Maximização da cobertura capacitada	Este método escolhe instalações de modo a que a totalidade ou a maior parte da procura possa ser servida sem exceder a capacidade de qualquer instalação
	Minimizar instalações	Este método escolhe instalações de modo a que o maior número de pontos de procura possíveis esteja dentro do corte de impedância. Além disso, o número de instalações necessárias para cobrir todos os pontos de procura é minimizado
	Maximizar o atendimento	O método de maximizar o atendimento escolhe instalações de modo que seja alocado para as instalações o maior peso da procura, assumindo que o peso da procura diminui com a distância
	Maximizar a quota de mercado	O método de maximizar a quota de mercado escolhe instalações de modo que a maior quantidade de procura alocada seja capturada na presença de concorrentes.
	Capturar a quota de mercado	O método de capturar a quota de mercado funciona na presença de concorrentes e tenta escolher o menor número de instalações necessárias para capturar a quota de mercado que for especificada.

Em seguida, aplicaram-se as distintas funções objetivo para compreender o impacto que estas têm no tema em estudo. Colocou-se as localizações com maior aptidão (0,99) como as instalações candidatas, e as lojas da concorrência (Pingo Doce, Minipreço, LIDL, Amanhecer) e as lojas Meu Super, já existentes, como lojas concorrentes às possíveis novas localizações. Como procura pelos estabelecimentos colocaram-se os centroides das áreas urbanas de Portugal Continental, e como valor de impedância, colocou-se 5 minutos, que corresponde à distância que os clientes estão dispostos a percorrer para fazerem as suas compras.

5.4.1. Minimização da Impedância (p-mediana)

Dos modelos localização-alocação o modelo de minimização da impedância (ou modelo p-mediana) é o mais simples, uma vez que tem como único objetivo minimizar as distâncias entre as possíveis lojas e a procura por estas.

Este modelo tem várias aplicações nos setores públicos e privados, mas sendo o objetivo do setor privado a minimização dos custos e a maximização da eficiência, este modelo reduzirá os custos gerais de transporte, tornando-se o modelo apropriado para o setor privado (Algharib, 2011). Quanto menor o tempo de deslocação dos clientes às lojas mais provável será a sua deslocação até às mesmas.

Assim sendo, este modelo tem como único objetivo a eficiência espacial, tentando encontrar as localizações que minimizam as distâncias percorridas pelos clientes. Ou seja, determinar que localizações minimizam o valor da função F definida da seguinte forma (Ramirez & Bosque, 2001):

$$F = \sum_i \sum_j D_{ei} T_{ij} X_{ij} \quad (5.5)$$

Em que F é a função de otimização; D_{ei} é a procura no ponto i (quantidade de utilizadores, habitantes, pessoas que precisam do serviço); T_{ij} é o custo do transporte desde o ponto de procura i ao ponto de oferta j . Normalmente emprega-se a distância entre os dois pontos; X_{ij} é um termo constante, a que se atribui valor 1 se o ponto de procura i é servido pelo ponto de oferta j , caso contrário atribui-se 0.

Assim, foi determinada a impedância em 5 minutos, e feito o cálculo para cada possível localização. O resultado é apresentado na Figura 5.14, sendo que os pontos de procura fora do limite de impedância não serão alocados e um ponto que esteja dentro do limite de duas ou mais instalações, apenas é alocado para a instalação mais próxima.



Figura 5.14: Resultado do modelo de minimização da impedância

De acordo com os resultados obtidos as quatro localizações que alocam maior procura na menor distância possível (dentro do limite da impedância) estão localizadas no distrito de Lisboa.

A localização que aloca mais população está na união de freguesias de Moscavide e Portela, abastecendo cerca de 4 514 residentes. A localização foi escolhida em detrimento de outra, na mesma área, pois esta alocava mais residentes no limite da impedância. Na área envolvente encontram-se duas lojas concorrentes, a 2 e 3 minutos da loja escolhida que não tiveram influência neste modelo.

A segunda localização mais atrativa está localizada na freguesia de Benfica, e abastece cerca de 2 968 residentes. Esta localização foi escolhida em detrimento de outras

9 possíveis localizações, por ser a que alocava mais residentes. Na área envolvente encontram-se duas lojas concorrentes, a 3 e 4 minutos da loja escolhida.

A terceira localização que aloca mais população está situada na união de freguesias de Algés, Linda-a-Velha e Cruz Quebrada-Dafundo, conseguido alocar cerca de 2 811 residentes. Dentro do limite de impedância, encontra-se um concorrente a cerca de 1 minuto da loja escolhida, e outros dois a menos de 10 minutos.

Por último, a quarta localização situa-se na união de freguesias de Agualva e Mira-Sintra e abastece 2 210 residentes. Esta foi escolhida em detrimento de outra localização (mais próxima do concorrente), e dentro do limite de impedância existe uma loja concorrente a menos de 2 minutos da loja escolhida.

5.4.2. Maximizar a Cobertura

Este modelo, desenvolvido por Church e Reville (1974), procura encontrar os locais ideais para um número fixo de instalações que cubra o maior número possível de pontos de procura dentro do limite de impedância. Ou seja, procura maximizar o número total de pontos de procura cobertos dentro de um limite de distância ou tempo de serviço (R), localizando as instalações candidatas próximas à densidade populacional. Isto significa que a instalação localizada onde a densidade de procura é elevada tem a vantagem de ser seleccionada primeiro.

Este modelo é expresso da seguinte forma (Buzai, 2013):

$$F = \sum_{i \in I} a_i x_i \quad (5.6)$$

Onde I é o grupo de pontos de procura (indexado por i); a_i é a população no nó de procura i ; e x_i é 1 se o centro de procura i está localizado dentro da área de cobertura ($x_i \leq R$), e 0 caso contrário.

Definidos os parâmetros necessários para a área em estudo, o resultado é apresentado na Figura 5.15.



Figura 5.15: Resultado do modelo de maximização da cobertura

À semelhança do modelo de minimização da impedância, as quatro localizações que maximizam a cobertura, dentro do limite da impedância, estão no distrito de Lisboa, mais precisamente na união de freguesias de Moscavide e Portela, na freguesia de Benfica, na união de freguesias de Algés, Linda-a-Velha e Cruz Quebrada-Dafundo, e na união de freguesias de Aqualva e Mira-Sintra.

Este modelo apresenta, exatamente, os mesmos resultados que o modelo anterior, o que quer dizer que as localizações que permitem a maximização da cobertura são também as que minimizam a impedância.

5.4.3. Minimizar as Instalações

O modelo de minimização de instalações seleciona o número de instalações necessárias para todos os pontos de procura, ou os que sejam possíveis, que estejam dentro dos limites de impedância. Ou seja, este modelo vai maximizar a quantidade de procura que pode ser abrangida dentro do número mínimo de instalações e da distância ou tempo especificado.

Este modelo é similar ao de maximização da cobertura, com a exceção do número de instalações a localizar, que neste caso é determinado pelo algoritmo ao invés do utilizador. Assim, o algoritmo tenta maximizar a quantidade de pontos de procura, minimizando o número de instalações.

O resultado deste modelo é apresentado na Figura 5.16. Para este modelo utilizou-se as lojas Meu Super já existentes, como lojas escolhidas, e juntaram-se as localizações possíveis como localizações candidatas. É de ter em consideração que neste modelo ainda não é tido em consideração a influência das lojas concorrentes existentes e a resposta por parte destas.

De acordo com os resultados, das 215 localizações possíveis, apenas 106 destas foram escolhidas pelo modelo, e das 123 lojas MS existentes, apenas 87 foram escolhidas. Ou seja, apenas 87 localizações alocam o máximo de clientes numa área de impedância de 5 minutos.

Das lojas Meu Super já existentes, as lojas que mais alocam clientes são o MS de Campo de Ourique, com 3 071 residentes, o MS da Reboleira, que atrai 2 430 residentes, o MS de Moscavide, que aloca 2 272 residentes, e o MS de Vila Chã, na freguesia de Mira de Água (Amadora), com 1 880 residentes. Houve lojas MS não utilizadas pelo modelo em 12 distritos de Portugal Continental (com exceção de Coimbra, Évora, Faro, Guarda, Viana do Castelo e Vila Real), das quais 12 delas estão localizadas no distrito de Lisboa, 6 no Porto, e 4 em Braga e Viseu.

A localização que aloca mais população está situada na união de freguesias de Algés, Linda-a-Velha e Cruz Quebrada-Dafundo, conseguindo alocar cerca de 2 811 residentes. Em seguida, a localizada na união de freguesias de Moscavide e Portela, abastecendo cerca de 2498 residentes, e duas localizações na freguesia de Lisboa, conseguindo alocar 1 897 e 1 798 potenciais clientes.

As localizações que alocam menos população estão localizadas nas freguesias de Setúbal (São Sebastião), em Santa Maria, São Pedro e Matacães, Pombalinho, e na União das freguesias de Santarém (Marvila), Santa Iria da Ribeira de Santarém, Santarém (São Salvador) e Santarém (São Nicolau). Estas localizações apenas conseguem alocar o máximo de 8 potenciais clientes.

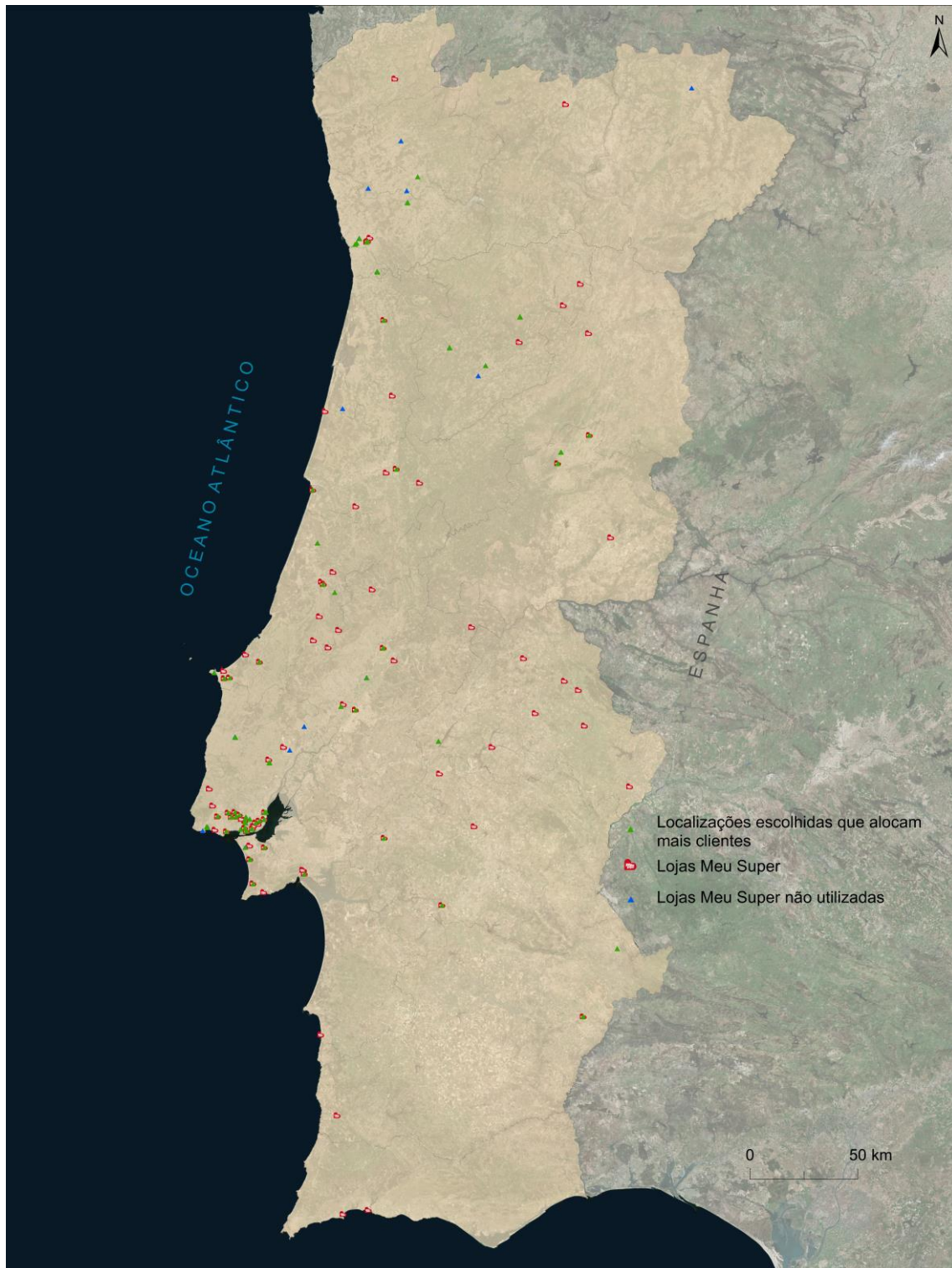


Figura 5.16: Resultado do modelo de minimização das instalações

5.4.4. Maximizar o Atendimento

O modelo de maximização do atendimento, desenvolvido por Holmes (1972), tem como objetivo determinar as localizações que maximizam o atendimento às instalações.

Este modelo procura maximizar a quantidade de pontos de procura que a instalação pode cobrir ou atender, com um limite de impedância especificado. O modelo de maximização de atendimento pressupõe que a probabilidade de interação entre as localizações das instalações e os locais de procura diminui com uma distância crescente (a probabilidade de atendimento diminui linearmente quando a distância aumenta). Portanto, este modelo tende a localizar instalações de modo a que elas estejam próximas do maior número de pontos de procura.

Assim, a Figura 5.17 mostra o resultado deste modelo aplicado aos dados desta dissertação. É de relembrar que apesar de se terem colocado as lojas concorrentes, estas não entram para o cálculo do modelo.



Figura 5.17: Resultado do modelo de maximização do atendimento

Semelhantemente aos restantes modelos é no distrito de Lisboa que se encontram as quatro localizações que alocam maior procura na menor distância possível (dentro do limite da impedância). No limite da impedância, a procura total equivale a 5050 residentes (53 pontos de procura).

A localização que aloca mais população está localizada na união de freguesias de Algés, Linda-a-Velha e Cruz Quebrada-Dafundo, abastecendo cerca de 1433 residentes. Esta localização foi escolhida em detrimento de outras 9 possíveis localizações, por ser a que aloca mais residentes. Na área envolvente encontram-se duas lojas concorrentes, a 3 e 4 minutos da loja escolhida, que não tiveram influência neste modelo.

A segunda localização mais atrativa está situada na união de freguesias de Moscavide e Portela, e abastece cerca de 1333 residentes. A localização foi escolhida em detrimento de outra, na mesma área, pois esta aloca mais residentes no limite da impedância. Na área envolvente encontram-se duas lojas concorrentes, a 2 e 3 minutos da loja escolhida.

A terceira localização que aloca mais população está situada na freguesia de Marvila, conseguindo alocar cerca de 1170 residentes.

Por último, a quarta localização situa-se na freguesia de Benfica e abastece 1113 residentes. Esta foi escolhida em detrimento de outra localização (mais próxima do concorrente), e dentro do limite de impedância existe uma loja concorrente a menos de 3 minutos da loja escolhida.

5.4.5. Maximizar a Quota de Mercado

Devido à quantidade de dados recolhidos foi possível utilizar os modelos localização-alocação com a componente da concorrência. O sector da venda a retalho é extremamente complexo, por isso é expectável que um modelo exato do comportamento dos consumidores ou da resposta da competição também seja complexo. Revelle (1986) apresentou os modelos localização-alocação à localização competitiva. E Goodchild (1984) sugeriu o Modelo *Market Share* dentro dos modelos localização-alocação.

O Modelo *Market Share* (MMS) é uma estratégia conservadora de evasão, para encontrar lacunas na cobertura da concorrência no mercado. Dado um conjunto de pontos representantes das lojas em estudo e um conjunto de pontos dos concorrentes da insígnia em estudo, este modelo vai selecionar as localizações que capturem maior quantidade de clientes, tendo em consideração a presença de concorrentes. Goodchild (1984) define a função do Modelo *Market Share* da seguinte forma:

$$MMS = \text{Max} \sum_i \sum_{j \in S \cup F} a_{ij} w_i \quad (5.7)$$

Onde a_{ij} é igual a 1 se $d_{ij} < d_{ik}$ e $K \in T$ (assumindo que não há laços), $K \neq j$, caso contrário, o valor de a_{ij} é 0.

Estando a aplicar estes modelos ao sector privado, este modelo não tem em consideração a resposta por parte dos concorrentes, quer a reação quer a antecipação da seleção das localizações para a marca em estudo, e assume que a concorrência se mantém fixa. Deste modo, o resultado é apresentado na Figura 5.18.



Figura 5.18: Resultado do modelo de maximização da quota de mercado

De acordo com os resultados obtidos é também em Lisboa que se situam as quatro localizações que alocam maior procura, dentro do limite da impedância, mesmo na presença de concorrentes. Visto que neste modelo tem-se a concorrência em consideração, assume-se que existe partilha dos clientes entre todos os agentes de mercado.

A localização que aloca mais população está localizada na união de freguesias de Moscavide e Portela, abastecendo cerca de 2498 residentes. Esta localização é mais vantajosa do que a localização dos concorrentes que estão dentro do limite da impedância, visto que os concorrentes apenas conseguem alocar 2066 (a 3 minutos da loja escolhida) e 1780 residentes (a 4 minutos).

A segunda localização mais atrativa está localizada na freguesia de Benfica, e abastece cerca de 1811 residentes. A 2 minutos da localização escolhida encontram-se dois concorrentes que conseguem abastecer 967 e 839 residentes, sendo que a loja mais próxima da escolhida é a que apresenta maior alocação da procura, entre os concorrentes.

A terceira localização que aloca mais população está situada na união de freguesias de Massamá e Monte Abraão e consegue abastecer cerca de 1480 residentes. Neste caso, o concorrente que está dentro do limite de impedância (a menos de 2 minutos da loja escolhida), aloca mais residentes do que a localização escolhida, cerca de 1789 residentes.

Por último, a quarta localização situa-se na freguesia de Benfica e abastece cerca de 1471 residentes. Esta localização, segundo o modelo, apenas partilha os clientes de um quarteirão com um concorrente que se encontra fora do limite de impedância.

5.4.6. Capturar a Quota de Mercado

O modelo de captura de mercado alvo determina o número mínimo de instalações necessárias para capturar uma percentagem específica do total de quota de mercado, apesar da existência de concorrentes. A quota de mercado total equivale à soma de todas as quotas de todos os agentes, num determinado mercado, sendo assim necessário saber os pesos (quotas) das instalações em estudo e das instalações concorrentes.

Assim, antes de realizar este modelo, foi necessário saber a percentagem de mercado que as lojas Meu Super detêm no território em estudo. No entanto, essa informação não se encontrava disponível, consequentemente, procurou-se o valor de *market share* que a Sonae tem no sector de retalho alimentar, em Portugal. Encontram-se dois valores, o primeiro correspondente a 2013, em que a Sonae apresentava o *market share* de 25,9%, e o segundo, de 2017, em que a Sonae apenas representava 21,9% do mercado.

Assim realizaram-se dois modelos, com os dois valores. Aquando finalização dos modelos, este encontrou o valor de 0.89 como o valor de otimização, fazendo com que qualquer valor que se coloque não produza alterações no modelo. Assim sendo, os resultados obtidos com 0.89 de *target market* são demonstrados na Figura 5.19.

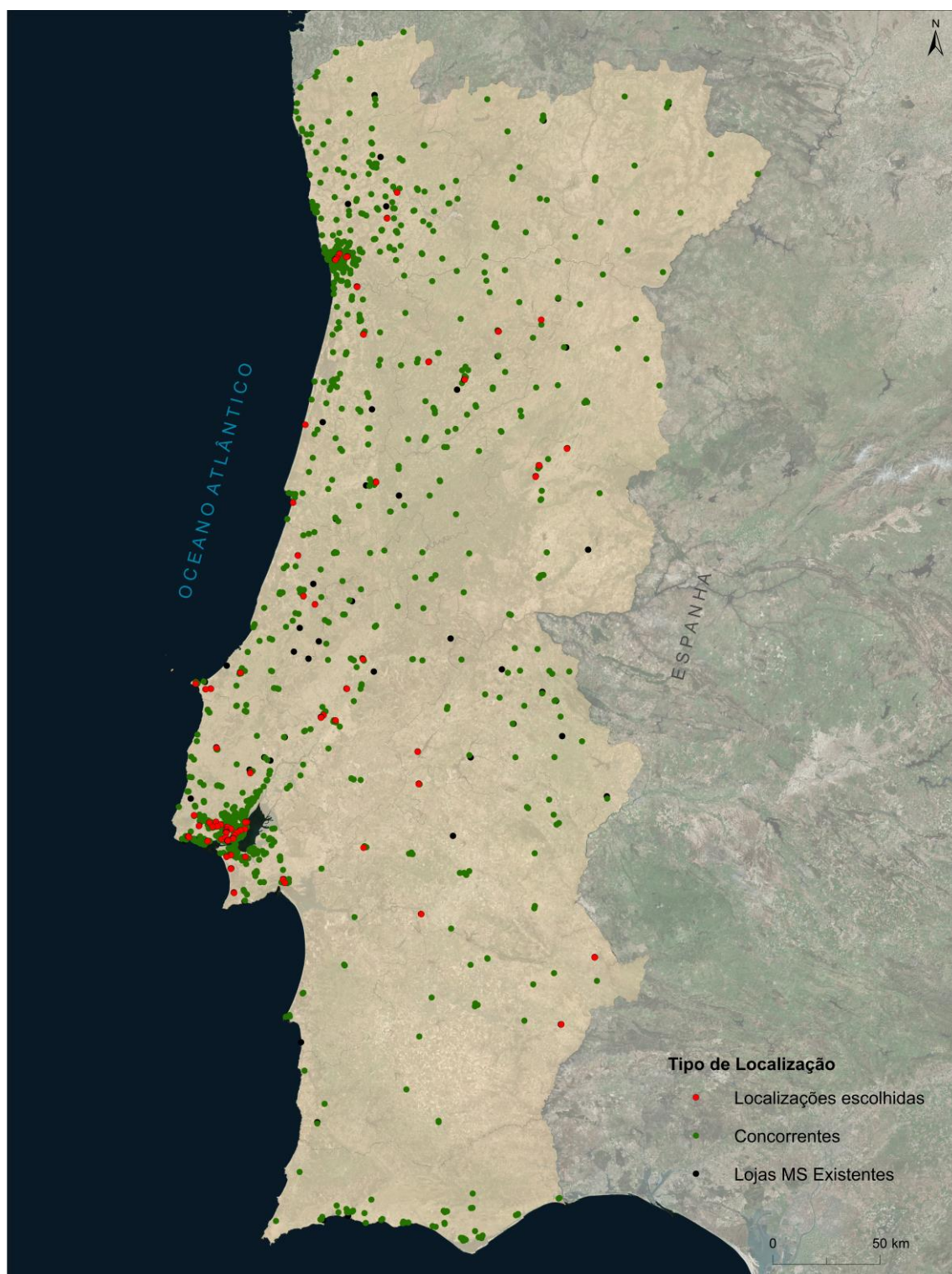


Figura 5.19: Resultado do modelo de Target Market Share, referente ao ano de 2017

De acordo com os resultados obtidos, 1 125 pontos correspondem a lojas concorrentes à insígnia em estudo, 123 às lojas Meu Super já existentes, e as restantes 118 são as localizações possíveis escolhidas para que as lojas MS conseguissem 0,89% do mercado existente.

Das lojas Meu Super já existentes, as que mais alocam clientes são o MS de Campo de Ourique, com 2 231 residentes, o MS de Moscavide, que aloca 1 880 residentes, e o MS de Vila Chã, na freguesia de Mira de Água (Amadora), com 1 809 residentes. As que alocam menos população, aproximadamente um residente, são os MS de Golegã, Póvoa de Santarém e Aroeira.

Das possíveis localizações são as localizadas na união de freguesias de Moscavide e Portela (1 975 e 1 534 residentes), na união de freguesias de Massamá e Monte Abraão (1 480 residentes), na freguesia de Benfica (1 470 residentes) e na união de freguesias de Agualva e Mira-Sintra (1 282 residentes). Já as localizações que atraem menos população, sensivelmente um residente, estão localizadas nas freguesias de Pombalinho, Santarém (Marvila) e Santa Iria da Ribeira de Santarém.

Relativamente aos concorrentes, a insígnia com mais alocação de residentes é o Minipreço, com as quatro lojas com maior alocação de residentes, nomeadamente as lojas Minipreço de Linda-a-Velha (3 459 residentes), Brandoa (3 014 residentes), na freguesia da Misericórdia (2 923 residentes) e na Tapada das Mercês (2 909 residentes). Por comparação, as lojas que alocam menor população, aproximadamente um residente, são da insígnia Pingo Doce, nomeadamente as de Póvoa de Santo Adrião, da Quarteira e de Amarante.

6. Considerações finais

Atualmente, as pessoas tendem a satisfazer uma boa parte das suas necessidades de consumo na proximidade do seu local de trabalho ou residência. Assim, houve por parte dos grandes grupos retalhistas um maior aproveitamento da proximidade aos clientes através dos estabelecimentos comerciais de proximidade. Através de contratos de *franchising* os grandes grupos retalhistas conseguem partilhar o risco de exploração com os pequenos retalhistas, podendo aproveitar os estabelecimentos já existentes ou criando estabelecimentos de raiz, moldando-os à imagem do grupo.

Devido às particularidades do setor da venda a retalho, a localização de equipamentos torna-se o fator essencial para o sucesso de um negócio. Uma boa localização permite maximizar lucros e minimizar os custos, através da proximidade ao maior número de clientes. Deste modo, uma análise pormenorizada dos possíveis locais para a abertura de novos estabelecimentos comerciais de proximidade permitirá conhecer a viabilidade das localizações e saber qual das localizações permite potenciar os seus serviços e lucros.

Os algoritmos de modelação incluem métodos de identificação do “perfil de dados”, baseados em abordagens estatísticas simples que usam a distância (*e.g.*: euclidiana, *Manhattan*, *grower*, etc.) e valores de amostra conhecidos para classificar os dados; métodos de regressão como as *Multivariate adaptive regression splines* (MARS) já aplicadas no contexto do comércio por Afonso (2018); e os métodos de *Machine Learning* (aprendizagem de máquina) como as *Boosted regression trees* (BRT), também testadas por Afonso (2018) e as *Support vector machines* (SVM). Dentro do contexto destes últimos métodos, podemos ainda encontrar as abordagens baseadas em inteligência artificial como as redes neuronais artificiais e os algoritmos genéticos (*e.g.*: *Genetic algorithm for rule set production* [GARP]).

O modelo *Environmental distance* é bastante simples e de fácil representação/interpretação. Sendo linear não se sobreajusta à amostra e, portanto, apresenta valores de probabilidade para todas as localizações (*i.e.*, células) possíveis. Contudo, não tem em conta as correlações e interações potenciais entre as variáveis e atribui um peso igual a todos os preditores, logo não permite avaliar a influência de cada preditor no padrão de distribuição potencial das lojas.

Por sua vez, as *Support vector machine* (SVM) são um método de classificação conhecido por apresentar grande capacidade de generalização. As SVM podem produzir resultados de classificação precisos e robustos, com uma base teórica sólida, mesmo quando os dados de entrada não são monótonos e/ou linearmente separáveis.

Assim, estes modelos podem ajudar a analisar a informação mais relevante de forma conveniente. Como linearizam os dados de forma implícita por meio da transformação de *kernel*, a precisão dos resultados não depende da interpretação do analista para a escolha ideal da função de linearização de dados (não-lineares) de entrada. As SVM operam localmente, de modo que são capazes de refletir as características individuais das lojas, comparando as suas variáveis de entrada com as das lojas na amostra de treino.

Embora as SVM não forneçam uma função paramétrica, a sua aproximação linear local pode providenciar um suporte importante para reconhecer os mecanismos que ligam diferentes variáveis à classificação final de uma loja. Por estas razões, as SVM são consideradas uma ferramenta útil e efetiva, de complemento às informações obtidas através dos modelos tradicionais de classificação linear.

No entanto, as SVM também possuem uma vulnerabilidade que é a dificuldade em determinar os valores ideais dos seus parâmetros. As SVM calculam o melhor separador linear no espaço de entrada de acordo com os dados de amostra. Para classificar dados que não são separáveis linearmente, o SVM usa funções de *kernel* (e.g.: linear, polinomial, função base radial (RBF) e sigmoide) para transformar os dados em elementos separáveis linearmente num espaço de dados de dimensão mais elevada. Para resolver o problema, tem sido proposta a utilização de algoritmos genéticos para busca dos valores de parâmetros ótimos, aumentando assim a precisão da predição.

O GARP é um modelo não-linear e não-paramétrico que não é influenciado pela multicolinearidade. É um método local eficiente que pretende mimetizar o processo de evolução natural postulado por Darwin. Deste modo é estabelecido um processo iterativo de otimização, incorporando a competição entre lojas pelo espaço, até se atingir a solução ótima. As principais desvantagens deste modelo são a sua complexidade (i.e., dificuldade de implementação), o tempo de processamento e o facto de possibilitar uma análise da importância das variáveis.

Apesar do openModeller ser mais utilizado para fenómenos naturais, os modelos de distribuição potencial apresentaram bons desempenhos, conseguindo ilustrar não somente variação na extensão das áreas consideradas aptas, como também mudanças nas áreas entre os diferentes resultados. Resultando num modelo síntese com variações nas áreas aptas de acordo com a influência dos três modelos, permitindo assim retirar as localizações com melhor aptidão para as novas lojas MS.

Os modelos de localização são considerados computacionalmente complexos e morosos. Além da tarefa computacional envolvida na solução de um modelo de localização, há várias questões importantes relacionadas especificamente com o modelo de integração e os SIG, nomeadamente, a compatibilidade dos dados, a exatidão na localização das instalações e da procura por estas, a agregação dos dados, os erros que podem existir na modelação, e a visualização dos resultados.

Neste sentido, os modelos localização demonstraram ser uma ferramenta essencial para encontrar as localizações ótimas para a abertura de novos estabelecimentos comerciais de proximidade. Apesar de morosos, estes modelos permitiram a introdução de uma grande quantidade de dados, em diferentes geometrias (linhas, pontos, polígonos) e origens (INE, DGT, Nokia), resultando num ficheiro em forma de ‘teia-de-aranha’ que permite uma visualização fácil de onde as lojas estão localizadas, quais os pontos de procura atraídos para cada loja e as diferenças potenciais nas áreas de serviço.

Com a conclusão da dissertação e concretização dos modelos de localização, chegou o momento de referir algumas considerações. De acordo com os resultados dos modelos localização-alocação, as melhores localizações para a abertura de novos estabelecimentos comerciais de proximidade Meu Super estão localizadas no distrito de Lisboa, mais precisamente na Grande Lisboa (freguesias de Moscavide e Portela, Benfica, Algés, Linda-a-Velha, Cruz Quebrada e Dafundo). Aqui, as localizações conseguem minimizar o custo de deslocação ao maior número de pessoas, e ainda são as localizações que alocam mais população, mesmo tendo em consideração as lojas concorrentes existentes na área.

Com estes resultados, e apesar da pequena quantidade de dados disponibilizados sobre as lojas Meu Super existentes, foi possível observar através dos modelos *target market share* (captura da quota de mercado) que as lojas MS que o modelo considerou que alocavam mais população, nomeadamente Campo de Ourique, Moscavide e Vila Chã,

estão entre as lojas que apresentam maior volume de vendas por m². O que faz com que este modelo não esteja muito longe da realidade, no que se refere às lojas com maior alocação de população. No entanto, quando se analisa as lojas que alocavam menor número de população, nomeadamente os MS de Aroeira, Golegã e Póvoa de Santarém, o mesmo não acontece. Para estas lojas, apenas a loja da Golegã apresenta valores similares aos resultados do modelo, as restantes ainda apresentam valores elevados de volume de vendas por m².

Quando se analisa a quantidade de transações comerciais, que realmente nos podem dar uma ideia em relação à quantidade de clientes que visitam as lojas, e os modelos de localização é possível observar que a loja de Moscavide é, sem sombra para dúvidas, a loja que apresenta melhores resultados, quer em termos localização e alocação, quer em termos de vendas e transações. Em relação às restantes lojas, nem as consideradas as melhores localizações, nem as consideradas piores, são na realidade as que apresentam os maiores ou menores valores de transações, *e.g.*, o MS de Póvoa de Santarém que, de acordo com o modelo, tem uma das piores localizações, uma das lojas que tem mais transações comerciais realizadas.

Apesar de algumas discrepâncias entre os modelos e a realidade, os resultados não estão incorretos. Apenas vêm provar que apesar de a localização ser importante, os dados empregues nos modelos podem ser sempre mais e melhores, e que existem outros fatores que influenciam o processo de decisão da localização de novas lojas e a escolha dos consumidores por determinados estabelecimentos comerciais que podem explicar os valores de vendas e transações aí realizadas (*e.g.*: a presença de outros serviços perto de certos estabelecimentos, a política de preços praticados, a preferência dos consumidores por determinada insígnia em detrimento das outras ou simplesmente o trajeto que os clientes preferem percorrer).

No fundo, todos os modelos têm vantagens e limitações. No entanto, estes são apenas uma aproximação (melhor ou pior) da realidade, podendo, por vezes, conter erros, mas mesmo assim pode tornar-se uma ferramenta útil para estudos futuros.

O progresso desta dissertação não aconteceu sem algumas dificuldades e percalços. Sentiu-se a primeira dificuldade com os dados provenientes da Sonae, que se revelaram reduzidos para se realizar uma caracterização fiável das lojas existentes. Uma vez que se tivesse uma quantidade de dados mais substancial os resultados poderiam ser

mais consistentes com a realidade. Também não foi possível obter informação concreta sobre o estilo de vida da população residente, o que poderia ter trazido mais pormenor e exatidão à caracterização da procura.

No futuro seria interessante utilizar outro *software* e tecer comparações entre os modelos existentes e os resultados obtidos nos dois *softwares*. Era importante também obter mais dados referentes às lojas Meu Super em funcionamento, assim como dados sobre os clientes que utilizam estas lojas, e ver como a junção de mais dados iria influenciar a escolha das novas localizações para estabelecimentos comerciais de proximidade. Por outro lado, seria interessante estudar a razão pela qual as lojas MS que estão localizadas em áreas de aptidão elevada encerraram, visto que apesar de a proximidade a outras lojas MS poder ter sido uma das razões para o encerramento, terá havido outras razões.

7. Bibliografia

- Afonso, S. (2018). *Ótimização das insígnias satélite da Sonae MC: Caso estudo Meu Super*. Dissertação de Mestrado em Sistemas de Informação Geográfica e Modelação Territorial aplicados ao Ordenamento para obtenção do grau de Mestre. Instituto de Geografia e Ordenamento do Território. Universidade de Lisboa.
- Algharib, S. M. (2011). *Distance and coverage: an assessment of location-allocation models for fire stations in Kuwait City, Kuwait*. Kent State University. Obtido de http://rave.ohiolink.edu/etdc/view?acc_num=kent1310950444
- American Marketing Association. (2013). Definition of Marketing. Consultado a 9 de Janeiro de 2018, em <https://www.ama.org/AboutAMA/Pages/Definition-of-Marketing.aspx>
- APED. (2012). Balanço de Recursos Humanos. *Associação Portuguesa de Empresas de Distribuição*, 62(10), 2–17. Obtido de http://aped.pt/application/files/1114/5406/4149/393_1_G.pdf
- Applebaum, W. (1951). Studying customer behavior in retail stores. *The Journal of Marketing*, 16(2), 172–179. <https://doi.org/10.2307/1247625>
- Applebaum, W. (1966). Methods for determining Store Trade Areas, Market Penetration, and Potential Sales. *Journal of Marketing Research*, 3(2), 127–141. <https://doi.org/10.1017/CBO9781107415324.004>
- Aranha, F., & Figoli, S. (2001). Geomarketing : Memórias de Viagem. Em *GeoBrasil* (pp. 1–73). (Versão Preliminar). São Paulo.
- Aronoff, S. (1989). *Geographic information systems: a management perspective*. (S. Aronoff, Ed.). Michigan: WDL Publications. <https://doi.org/10.1080/10106048909354237>
- Baray, J., & Cliquet, G. (2003). Retail Store Location: An application of signal processing to the detection of location opportunities for bakeries in France. Em *Fourth International Conference on Operations and Quantitative Management (ICOQM IV)* (pp. 1–13). Taiwan: National Cheng Kung University Tainan. Obtido de <http://membres.multimania.fr/geomarketing/articles/LocationOpportunitiesDetection.pdf>
- Barreto, F. (2008). *Modelagem de distribuição potencial de espécies como ferramenta para a conservação, seleção e avaliação de algoritmos e aplicação com Heliconius nattereri Felder, 1865 (NYMPHALIDAE: HELICONIINAE)*. Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós- Graduação em Entomologia, para obtenção do título de Doctor Scientiae. Universidade Federal de Viçosa. Obtido de http://www.tede.ufv.br/tedesimplificado/tde_busca/processaArquivo.php?codArquivo=1952
- Baviera-Puig, A., Buitrago-Vera, J., & Escriba-Perez, C. (2016). Geomarketing models in supermarket location strategies. *Journal of Business Economics and Management*, 17(6), 1205–1221. <https://doi.org/10.3846/16111699.2015.1113198>
- Benévolo, L. (1993). *The European City*. Oxford: Wiley-Blackwell.

- Berman, B., & Evans, J. (1995). *Retail Management*. (B. Berman & J. Evans, Eds.) (6ª Edição). Harlow, United Kingdom: Prentice-Hall.
- Bernardes, E. B. (2017). *Geomarketing: Utilização De Um Sistema De Informação Geográfica (Sig) Para a Análise Da Distribuição E Padrões Dos Alunos Ingressantes na Faculdade Redentor de Itaperuna*. Monografia apresentada a Faculdade Redentor como parte dos requisitos para a obtenção do título de bacharel em Engenharia de Produção. Faculdade Redentor. Obtido de <http://www.interscienceplace.org/isp/index.php/isp/article/view/579>
- Berry, M., & Linoff, G. S. (1996). *Data mining techniques* (Second Edi). Indiana: Wiley Publishing. <https://doi.org/http://doi.acm.org/10.1145/235968.280351>
- Birkin, M., Clarke, G., & Clarke, M. (2002). *Retail Geography & Intelligent Network Planning*. Chichester, England: John Wiley and Sons.
- Boudot, J.-D. (1999). *Geomarketing*. Ecole Nationale Des Sciences Geographiques, Marne-La-Vallée.
- Bradford, M. G., & Kent, W. A. (1987). Utilização do Solo Agrícola: O Modelo de Thüinen. Em *Geografia Humana. Teoria e suas aplicações* (Gradiva, pp. 46–66). Lisboa.
- Britto, D., Filho, F., & Alexandre, J. (2009). Desvendando os Mistérios do Coeficiente de Correlação de Pearson. *Revista Política Hoje*, 18(1), 115–146.
- Burrough, P. A. (1986). Principles of geographical information systems for land resources assessment. *Geocarto International*, 1(3), 54. Obtido de <http://www.scopus.com/inward/record.url?eid=2-s2.0-0022850320%7B&%7DpartnerID=tZOtx3y1>
- Burton, E., & Mitchell, L. (2006). *Inclusive Urban Design: Street for Life* (First Edit). Oxford: Elsevier. <https://doi.org/10.1017/CBO9781107415324.004>
- Buzai, G. (2013). Location – allocation models applied to urban public services . Spatial analysis of Primary Health Care Centers in the city of Luján , Argentina. *Hungarian Geographical Bulletin*, 62(4), 387–408.
- Cabral, A. (2012). *Análise Multicritério Em Sistemas De Informação Geográfica Para a Localização De Aterros Sanitários*. Dissertação de Mestrado em Gestão do Território, Área de Especialização em Detecção Remota e Sistemas de Informação Geográfica, para obtenção do grau de Mestre. Faculdade de Ciências Sociais e Humana. Universidade Nova de Lisboa.
- Cachinho, H. A. P. (1999). *O Comércio Retalhista Português na (Pós-)modernidade*. Dissertação de Doutoramento para a obtenção do grau de Doctor em Geografia Humana. Faculdade de Letras. Universidade de Lisboa.
- Canhos, V. P. (2005). *openModeller: a framework for species distribution modeling. Thematic Research Project - Suported by FAPESP*. São Paulo, Brasil. Obtido de http://openmodeller.cria.org.br/documentos/relatorios/openModeller_Report_4.pdf
- Cardoso, C. E. P. (2011). *Geomarketing como Suporte de Decisão em Gestão do Território*. Dissertação de Mestrado em Gestão do Território na Área de Especialização em Detecção Remota e Sistemas de Informação Geográfica, para obtenção do grau de Mestre. Faculdade de Ciências Sociais e Humanas.

Universidade Nova de Lisboa.

- Castro, A. C. B. (2007). *As vendas e o espaço no retalho: Modelos Económétricos aplicados a um grupo de distribuição alimentar potuguês*. Tese de Mestrado em Ciências Empresariais, para obtenção do grau de Mestre. Faculdade de Economia. Universidade do Porto.
- Catunda, R. L. (2010). *Geomarketing: Uma Análise Da Distribuição Geográfica Das Instituições De Ensino Superior*. Mestrado Profissional em Administração e Controladoria, para obtenção do grau de Mestre. Faculdade de Economia, Administração, Atuária e Contabilidade. Universidade Federal do Ceará.
- Cavion, R., & Philips, J. (2006). Os Fundamentos do Geomarketing: Cartografia, Geografia e Marketing. Em R. Cavion & J. Philips (Eds.), *Congresso Brasileiro de Cadastro Técnico Multifinalitário* (pp. 1–9). Florianópolis: UFSC.
- Cheng, E., Li, H., & Yu, L. (2007). A GIS approach to shopping mall location selection. *Building and Environment*, 42(2), 884–892.
- Church, R. L. (2002). Geographical information systems and location science. *Computers and Operations Research*, 29(6), 541–562. [https://doi.org/10.1016/S0305-0548\(99\)00104-5](https://doi.org/10.1016/S0305-0548(99)00104-5)
- Clarke, G. (1998). Changing methods of location planning for retail companies. *GeoJournal*, 45(4), 289–298. <https://doi.org/10.1023/A:1006995106736>
- Cliquet, G. (2006). *Geomarketing: Methods and strategies in spacial marketing*. (G. Cliquet, Ed.). Chippingham, Wiltshire: Wiley, ISTE ltd.
- Comber, A., Dickie, J., Jarvis, C., Phillips, M., & Tansey, K. (2015). Locating bioenergy facilities using a modified GIS-based location-allocation- algorithm: considering the spatial distribution of resource supply. *Applied Energy*, 154, 309–316. <https://doi.org/10.1016/j.apenergy.2015.04.128>
- Copeland, M. T. (1923). Relation of consumers' buying habits to marketing methods. *Harvard Business Review*, 1(3), 282–289.
- Costa, C. M. P. (2010). *Localização Óptima do Futuro Hospital de Sintra: Aplicação de modelos de location-allocation no planeamento de cuidados de saúde*. Dissertação de Mestrado em Ciência e Sistemas de Informação Geográfica, para obtenção do grau de Mestre. Instituto Superior de Estatística e Gestão da Informação. Universidade Nova de Lisboa.
- Cowen, D. J. (1988). GIS versus CAD versus DBMS : What Are the Differences ? *Engineering*, 54(11), 1551–1555.
- CRIA. (sem data). Projetos de Sistemas de Informação. Consultado a 22 de Abril de 2018, em <http://www.cria.org.br/projetos>
- Curry, D. (1993). *The New Marketing Research Systems: How to Use Strategic Database Information for Better Marketing Decisions*. New York: John Wiley & Sons, Inc.
- Davies, M., & Clark, I. (1994). A framework for network planning. *International Journal of Retail and Distribution Management*, 22(6), 6–10.
- Davies, R. L. (1976). The recent history and problems of small shops and related small businesses. Em P. Jones & R. Oliphant (Eds.), *Local shops: Problems and prospects*.

Reading: URPI.

- Dawson, J. (1995). Retail trends in Scotland: a review. *International Journal of Retail & Distribution Management*, 23(10), 4–20.
- Dawson, J. (2000). *Future Patterns of Retailing in Scotland*. The Scottish Executive Central Research Unit. Saxon House, Farnborough: Crown.
- Deuker, K. J. (1979). Land resource information systems: a review of fifteen years experience. *Geo-Processing*, 1(s.n), 105–128.
- Dias, P. (2013). *Análise exploratória de dados espaciais de actores de sectores estratégicos na cidade de Lisboa*. Relatório de Estágio de Mestrado em Gestão do Território, Área de Especialização em Detecção Remota e SIG, para a obtenção do grau de Mestre. Faculdade de Ciências Sociais e Humanas. Universidade Nova de Lisboa.
- Duch, N. (2005). *La teoría de la localización*. Tesis doctoral. Institut d’Economia de Barcelona. Universitat de Barcelona. Obtido de http://www.eco.ub.es/~nduch/postgrau_archivos/Duch_localizacion.pdf
- Eicher, C. L., & Brewer, C. A. (2001). Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*, 28(2), 125–138. <https://doi.org/10.1559/152304001782173727>
- ESRI. (sem data-a). An overview of fuzzy classes. Consultado a 17 de Julho de 2018, em <http://desktop.arcgis.com/en/arcmap/10.3/analyze/arcpy-spatial-analyst/an-overview-of-fuzzy-classes.htm>
- ESRI. (sem data-b). Average Nearest Neighbor. Consultado a 26 de Dezembro de 2017, em <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/average-nearest-neighbor.htm>
- ESRI. (sem data-c). Line Density. Consultado a 12 de Março de 2018, em <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-line-density-works.htm>
- ESRI. (sem data-d). Network elements. Consultado a 6 de Março de 2018, em <http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/network-elements.htm>
- ESRI. (sem data-e). What is a network dataset? Consultado a 6 de Março de 2018, em <http://desktop.arcgis.com/en/arcmap/latest/extensions/network-analyst/what-is-a-network-dataset.htm>
- ESRI. (2012). Location Analytics: The Next Big Step in Business Analysis. Consultado a 22 de Março de 2018, em <http://www.esri.com/news/arcnews/fall12/articles/location-analytics-the-next-big-step-in-business-analysis.html>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/10.1609/aimag.v17i3.1230>
- Fernandes, J., Cachino, H., & Ribeiro, C. (2000). *Comércio Tradicional em Contexto Urbano: Dinâmicas de Modernização e Políticas Públicas*. Gabinete de Estudos

de Desenvolvimento e Ordenamento do Território e Faculdade de Letras da universidade do Porto, Porto, Cardenos nº2.

- Figueiredo, R. (2014). *Áreas De Influência De Espaços Verdes Urbanos De Proximidade: Uma Abordagem Exploratória Na Freguesia De Arroios*. Dissertação para obtenção do Grau de Mestre em Urbanismo e Ordenamento do Território. Instituto Técnico. Universidade de Lisboa.
- Food Market Institute. (2018). Retail Segment Definitions. Consultado a 14 de Setembro de 2018, em <https://www.fmi.org/our-research/supermarket-facts>
- Furlan, A. (2011). Geoprocessamento: estudos de Geomarketing e as possibilidades de sua aplicação no planeamento do desenvolvimento socioeconómico. *GEOUSP: espaço e tempo*, 29(Especial), 97–105. <https://doi.org/http://dx.doi.org/10.11606/issn.2179-0892.geousp.2011.74208>
- Garcia, R. A. C. (2012). *Metodologias de Avaliação da Perigosidade e Risco associado a Movimentos de Vertente: Aplicação na bacia do rio Alenquer*. Dissertação de Doutoramento em Geografia Física, para a obtenção do grau de Doctor. Instituto de Geografia e Ordenamento do Território. Universidade de Lisboa.
- Garcia, R. A. C., Oliveira, S. C., & Zêzere, J. L. (2016). Assessing population exposure for landslide risk analysis using dasymetric cartography. *Natural Hazards and Earth System Sciences*, 16(12), 2769–2782. <https://doi.org/10.5194/nhess-16-2769-2016>
- Gerrard, R. A., & Church, R. L. (1996). Closest assignment constraints and location models: Properties and structure. *Location Science*, 4(4), 251–270. [https://doi.org/10.1016/S0966-8349\(97\)00001-6](https://doi.org/10.1016/S0966-8349(97)00001-6)
- Goldberg, D. W., Wilson, J. P., & Knoblock, C. a. (2007). From Text to Geographic Coordinates: The Current State of Geocoding. *URISA Journal*, 19, 33–46. <https://doi.org/10.1.1.119.714>
- González-Benito, Ó. (2002). Geodemographic and socioeconomic characterization of the retail attraction of leading hypermarket chains in Spain. *The International Review of Retail, Distribution and Consumer Research*, 12(1), 81–103. <https://doi.org/10.1080/09593960110103869>
- González-Benito, Ó., & González-Benito, J. (2005). The role of geodemographic segmentation in retail location strategy. *International Journal of Market Research*, 47(3), 295–316. <https://doi.org/Article>
- Goodchild, M. F. (1984). A location-allocation model for retail site selection. *Jornal of Retailing*, 60(1), 84–100.
- Goodchild, M. F. (2003). Geographic Information Science and Systems for Environmental Management. *Annual Review of Environment and Resources*, 28, 493–519. <https://doi.org/10.1146/annurev.energy.28.050302.105521>
- Guerreiro, C. (2014). *Mobilidade Pedonal e Comércio de Proximidade, que relação?* Dissertação de Mestrado em Urbanismo e Ordenamento do Território, para a obteção do grau de Mestre. Instituto Superior Técnico.
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2–3), 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)

- Guo, Q., Kelly, M., & Graham, C. H. (2005). Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling*, 182(1), 75–90. <https://doi.org/10.1016/j.ecolmodel.2004.07.012>
- Guo, Q., & Liu, Y. (2010). ModEco: An integrated software package for ecological niche modeling. *Ecography*, 33(4), 637–642. <https://doi.org/10.1111/j.1600-0587.2010.06416.x>
- Guptill, S. C., Cotter, D., Gilson, R., Liston, R., Tom, H., Trainor, T., ... Porter, E. (1988). *A Process for Evaluating Geographic Information Systems*. (S. C. Gupta, Ed.). Federal Interagency Coordinating Committee on Digital Cartography. U.S. Geological Survey.
- Haggett, P., & Chorley, R. J. (1969). *Network Analysis in Geography*. Somerset: Edward Arnold.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. (F. Bach, Ed.). Cambridge: MIT Press. <https://doi.org/10.2165/00002018-200730070-00010>
- Harris, R., & Longley, P. (2000). New Data and Approaches for Urban Analysis: Modelling Residential Densities. *Transactions in GIS*, 4(3), 217–234.
- Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and Neighbourhood Targeting*. West Sussex, England: John Wiley & Sons, Ltd.
- Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, 39(153), 41–57.
- Jones, K., & Simmons, J. (1990). *The Retail Environment*. London: Routledge.
- Jorge, V. (2009). José António Rousseau, Especialista em Distribuição. Consultado a 14 de Janeiro de 2018, em <http://www.hipersuper.pt/2009/12/17/quem-conduz-a-locomotiva-das-mudancas-sao-os-consumidores/>
- Jorge, V. (2011). Recheio Aposta no «Amanhecer». Consultado a 26 de Dezembro de 2018, em <http://www.hipersuper.pt/2011/01/27/recheio-aposta-no-amanhecer/>
- Júnior, A. (2002). *Fatores Influentes No Processo De Escolha Da Localização Agroindustrial No Paraná: Estudo De Caso De Uma Agroindústria De Aves*. Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título de Mestre em Engenharia de Produção. Universidade Federal de Santa Catarina.
- Kainz, W. (2001). *Fuzzy Logic and GIS*. Vienna, Austria: University of Vienna. Obtido de http://homepage.univie.ac.at/wolfgang.kainz/Lehrveranstaltungen/ESRI_Fuzzy_Logic/File_2_Kainz_Text.pdf
- Kemp, K. K. (2008). *Encyclopedia of geographic information science*. (K. K. Kemp, Ed.). Los Angeles: SAGE Publications.
- Kim, H., & Choi, B. (2013). The Influence of Customer Experience Quality on Customers' Behavioral Intentions. *Services Marketing Quarterly*, 34(4), 322–338. <https://doi.org/10.1080/15332969.2013.827068>
- Kotler, P. (1997). *Marketing management: analysis, planning, implementation and control*. *Journal of Marketing Management* (Vol. 7). Obtido de

<http://www.decitre.fr/pdf/feuilleter/9782744073458.pdf>

- Kotler, P. (2003). *Marketing insights from A to Z - 80 Concepts every manager needs to know*. Management. <https://doi.org/10.1017/CBO9781107415324.004>
- Kotler, P., & Armstrong, G. (2003). *Princípios de Marketing* (9ª edição). São Paulo: Pearson Prentice Hall.
- Latour, P., & Floc'h, J. (2001). *Géomarketing: principes, méthodes et applications*. Paris: Editions d'Organisation.
- Leszczyc, P., & Timmermans, H. (1997). Store-Switching Behavior. *Marketing Letters*, 8(2), 193–204. <https://doi.org/10.1023/A>
- Leventhal, B. (1993). Birds of a feather? Or, geodemographics – an endangered species? Em *Proceedings of the Market Research Society Conference* (pp. 223–239). Liverpool.
- Levine, N. (2002). *CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations, Part I*. The National Institute of Justice. Houston, Texas: U.S. Department of Justice.
- Levitt, T. (1960). *Marketing Myopia*. Boston, Massachusetts: Harvard Business Press.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic Information Systems and Science*. Hoboken: John Wiley & Sons, Inc.
- Longley, P., & Clarke, G. (1995). *GIS for Business and Service Planning*. GeoInformation International. Cambridge, MA: GeoInformation International.
- Lorena, L. A. N., Senne, E. L. F., Paiva, J. A. D. C., & Pereira, M. A. (2001). Integração de modelos de localização a sistemas de informações geográficas. *Gestão & Produção*, 8(2), 180–195. <https://doi.org/10.1590/S0104-530X2001000200006>
- Machado, C., Francisco, E., & Ribeiro, J. (2006). A Geodemografia e o Geomarketing na identificação de mercados potenciais. Em *Demografia dos negócios: Campo de estudo, perspectivas e aplicações* (pp. 77–98). Campinas: Associação Brasileira de Estudos Populacionais.
- Malczewski, J. (2004). GIS- based land use suitability analysis: a critical overview. *Progress in Planning*, 62(1), 3–65.
- Maroney, A. (1976). The decline of small shops in Liverpool and the problems involved in providing new facilities. Em *Local shops: Problems and prospects*. eds Jones, P and Oliphant, R. Reading: URPI.
- Martins, R. C. (2012). *Construção de um Modelo de Análise Espacial em SIG , que determine a Localização Ótima de Equipamentos Sociais para idosos , no Concelho de Lisboa*. Dissertação de Mestrado em Gestão do Território na Área de Especialização em Detecção Remota e Sistemas de Informação Geográfica, para obtenção do grau de Mestre. Faculdade de Ciências Sociais e Humanas. Universidade Nova de Lisboa.
- Mendes, A. B., & Themido, I. H. (2004). Multi Outlet Retail Site Location Assessment. *International Transactions in Operational Research*, 11(1), 32. Obtido de <https://repositorio.uac.pt/bitstream/10400.3/2143/1/ITOR04.pdf>

- Mennis, J. (2009). Dasymetric Mapping for Estimating Population in Small Areas. *Geography Compass*, 3(2), 727–745.
- Ministério do Comércio e Turismo. Decreto-Lei nº 339/85 de 21 de Agosto (1985). Governo da República.
- Motta, R. G., & Silva, A. V. da. (2006). Aumento da competição no varejo e seu impacto na indústria. *Revista Gerenciais*, 5(2), 101–108.
- Muñoz, M., Giovanni, R., Siqueira, M., Sutton, T., Brewer, P., Pereira, R. S., ... Canhos, V. P. (2011). openModeller: A generic approach to species' potential distribution modelling. *GeoInformatica*, 15(1), 111–135. <https://doi.org/10.1007/s10707-009-0090-7>
- Murad, A. A. (2015). Using GIS for Retail Location Assessment at Jeddah City. *International Journal of Advanced Computer Science and Applications*, 6(7), 125–134.
- Nielsen Market Track. (2012). *Scan Trends – Total Portugal*.
- Nielsen Market Track. (2013). *Scan Trends – Total Portugal*.
- Nielsen Market Track. (2014). *Scan Trends – Total Portugal*.
- Nielsen Market Track. (2015). *Scan Trends – Total Portugal*.
- Nielsen Market Track. (2017). *Scan Trends – Total Portugal*.
- Niti, D. (2007). *Retail Location Analysis: A Case Study of Burger King & McDonald's in Portage & Summit Counties, Ohio*. A thesis submitted to the College of Arts of Kent State University in partial fulfillment of the requirements for the degree of Masters of Arts. College of Arts of Kent State University. Obtido de http://rave.ohiolink.edu/etdc/view?acc_num=kent1196133312
- O'Malley, L., Patterson, M., & Evans, M. (1997). Retailer use of geodemographic and other data sources: an empirical investigation. *International Journal of Retail & Distribution Management*, 2(6), 188–196. <https://doi.org/10.1108/09590559710175935>
- Pacione, M. (2005). Urban Retailing. Em *Urban Geography: A Global Perspective* (Second Edi, pp. 326–358). New York: Routledge.
- Parente, J., & Kato, H. T. (2001). Área De Influência: Um Estudo No Varejo De Supermercados. *Revista de Administração de Empresas*, 41(2), 46–53. <https://doi.org/10.1590/S0034-75902001000200005>
- Parker, H. D. (1988). The Unique Qualities of a Geographic Information-System - A Commentary. *Photogrammetric Engineering And Remote Sensing*, 54(11), 1547–1549.
- PennState Eberly College of Science. (sem data). Detecting Multicollinearity Using Variance Inflation Factors. Consultado a 3 de Julho de 2018, em <https://onlinecourses.science.psu.edu/stat501/node/347/>
- Pettit, C., & Pullar, D. (1999). An integrated planning tool based upon multiple criteria evaluation of spatial information. *Computers, Environment and Urban Systems*, 23(5), 339–357.

- Polido, S. I. F. (2014). *Geocodificação Em Portugal: O Caso De Estudo Da Freguesia De Alfragide (Amadora)*. Relatório de Estágio de Mestrado em Sistemas de Informação Geográfica e Modelação Territorial Aplicados ao Ordenamento, para obtenção do grau de Mestre. Instituto de Geografia e Ordenamento do Território. Universidade de Lisboa.
- Puig, P. (2011). *The role of fuzzy logic in GIS modelling*. PhD Thesis for the degree of Doctor of Philosophy. Faculty of Engineering, Health , Science and the Environment. Charles Darwin University.
- PVANB. (2016). VIF stepwise variable selection. Consultado a 4 de Julho de 2018, em <https://pvanb.wordpress.com/author/pvanb/>
- RAI. (sem data). *Retailing & Retail Formats*. Syllabus from Rai Technology University. Dhodballapur Taluk, Bangalore.
- Ramirez, M., & Bosque, J. (2001). Localización de Hospitales : analogías y diferencias del uso del modelo P-mediano en SIG raster y vectorial. *Anales de Geografía de la Universidad Complutense de Madrid*, 21(1), 53–79. Obtido de <http://hum.unne.edu.ar/investigacion/geografia/labtig/publicaciones/public05.pdf>
- Revelle, C. (1986). The Maximum capture or sphere of influence problem: Hotelling revisited on a network. *Journal of Regional Science*, 26(2), 341–357.
- Rocha, J., & Henriques, C. (2014). A importância da análise espacial na gestão urbana. Um caso de estudo. *Interações*, 30(28), 21–30. Obtido de http://www.apgeo.pt/files/docs/CD_X_Coloquio_Iberico_Geografia/pdfs/096.pdf
- Rodeski, W. (2010). *Geomarketing: o uso da regressão logística múltipla no mapeamento de regiões geográficas de alto potencial mercadológico*. Monografia apresentada para a Universidade Federal do Rio Grande do Sul para a obtenção do grau de Bacharel em Estatística. Universidade Federal do Rio Grande do Sul.
- Roig-Tierno, N., Baviera-Puig, A., & Buitrago-Vera, J. (2013). Business opportunities analysis using GIS: the retail distribution sector. *Global Business Perspectives*, 1(3), 226–238. <https://doi.org/10.1007/s40196-013-0015-6>
- Roig-Tierno, N., Baviera-Puig, A., Buitrago-Vera, J., & Mas-Verdu, F. (2013). The retail site location decision process using GIS and the analytical hierarchy process. *Applied Geography*, 40(6), 191–198. <https://doi.org/10.1016/j.apgeog.2013.03.005>
- Rousseau, A. . (2008). *Manual de Distribuição* (2ª edição). Estoril: Princípia Editora.
- Salgueiro, T. B., & Cachinho, H. (2011). Retail Culture and Consumer Experiences. Em Barata-Salgueiro & Cachinho (Ed.), *Retail Planning for the Resilient City: consumption and urban regeneration* (pp. 147–157). Lisbon: Centro de Estudos Geográficos.
- Salgueiro, T. B., Cachinho, H., André, I. M., Leite, J., Bairrada, M., Rodrigues, M. J., & Feio, P. A. (2002). *Estratégias empresariais, emprego e empregabilidade no comércio*. Lisboa: Observatório do Comércio e Gabinete de Estudos e Prospectiva Económica.
- Sansone, M., & Colamatteo, A. (2017). Trends and Dynamics in Retail Industry : Focus on Relational Proximity. *International Business Research*, 10(2), 169–179. <https://doi.org/10.5539/ibr.v10n2p169>

- Santos, J. L. R. (2014). *Estratégias de Renovação do Retalho Alimentar Tradicional*. Dissertação para obtenção do Grau de Mestre em Engenharia e Gestão Industrial. Instituto Técnico de Lisboa.
- Santos, L. R. (2012). *Do hipermercado ao comércio de proximidade: o caso de estudo Amanhecer*. Projecto de Mestrado em Gestão, para obtenção do grau de Mestre. Instituto Superior de Ciências do Trabalho e da Empresa (ISCTE). Instituto Universitário de Lisboa.
- Schewe, C., & Smith, R. (1982). *Marketing: Conceitos, Casos e Aplicações*. São Paulo: Makron Books Mac Graw Hill, Brasil.
- Serafim, V. (2014). *Sistemas de Informação de Marketing, Geomarketing e Níveis Estratégicos em Organizações no Brasil*. Universidade Nova de Lisboa.
- Seven & i Holdings. (2012). *Convenience Store Operations*. Tokyo, Japan.
- Silva, F. (2009). *Um ensaio metodológico de cartografia dasimétrica aplicado à região Oeste e Vale do Tejo*. Dissertação apresentada à Faculdade de Letras da Universidade do Porto para a obtenção do grau de mestre em Sistemas de Informação Geográfica e Ordenamento do Território. Faculdade de Letras. Universidade do Porto. <https://doi.org/000190955>
- Silverman, B. W. (1986). ESRI: How Kernel Density works. Consultado a 26 de Fevereiro de 2018, em <http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/how-kernel-density-works.htm>
- Sleight, P. (1997). *Targeting Customers: How to Use Geodemographic and Lifestyle Data in Your Business*. Henley on Thames: NTC Publications.
- Sonae. (sem data). O Meu Super é Seu também. Consultado a 12 de Dezembro de 2017, em <https://www.meusuper.pt/pt/content/13-quem-somos>
- Sonae. (2015). *Sonae: Relatório de Contas*. Lisboa, Portugal.
- Stockwell, D., & Peters, D. (1999). The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, 13(2), 143–158. <https://doi.org/10.1080/136588199241391>
- Sui, D. Z. (1992). A fuzzy GIS modeling approach for urban land evaluation. *Computers, environment and urban systems*, 16(1), 101–115. [https://doi.org/10.1016/0198-9715\(92\)90022-J](https://doi.org/10.1016/0198-9715(92)90022-J)
- Teixeira, O. C. (1993). *SIG: sistemas de informação geográfica*. Dissertação em Engenharia Electrotécnica e de Computadores, com área de especialização de Informática Industrial, para obtenção do grau de Mestre. Departamento de Engenharia Electrotécnica e de Computadores. Universidade do Porto.
- The Economic Times. (sem data). Definition of Marketing Mix | What is Marketing Mix ? Consultado a 10 de Janeiro de 2018, em <https://economictimes.indiatimes.com/definition/marketing-mix>
- Townsend Peterson, A., Papeş, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography*, 30(4), 550–560. <https://doi.org/10.1111/j.2007.0906-7590.05102.x>
- VASP. (2017). Dicas para o Sucesso. *Quiosque*, 69(2), 1–32.

- Wells, L. (2018). Five key trends that will shape convenience retail in 2018. Consultado a 11 de Janeiro de 2019, em <https://www.talkingretail.com/opinion/five-key-trends-will-shape-convenience-retail-2018-16-01-2018/>
- Worboys, M. (1995). *GIS: A Computing Perspective*. Londres: Taylor and Francis.
- Worrall, L. (1991). *Spatial analysis and spatial policy using geographic information systems*. London: Belhaven Press.
- Yasnoff, W. A., & Miller, P. L. (2014). Decision Support and Expert Systems in Public Health. Em *Public Health Informatics and Information Systems* (2º Edition, pp. 449–467).
- Yrigoyen, C. (2003). El Geomarketing y la Distribución Comercial. *Revista Investigación y Marketing*, 79(s.n), 6–13.
- Zentes, J., Morschett, D., & Schramm-Klein, H. (2007). *Strategic Retail Management*. Gabler.
- Zentes, J., Morschett, D., & Schramm-Klein, H. (2011). Store Location – Trading Area Analysis and Site Selection. Em *Strategic Retail Management* (pp. 203–225). Wiesbaden: Gabler Verlag.
- Zulqarnain, H., Zafar, A. U., & Shahzad, M. (2015). Factors that affect the choice of Consumers in selecting Retail Store, for Grocery Shopping. *Internation Jornal of Multidisciplinary and Current research*, 3(12), 1167–1172. Obtido de <http://ijmcr.com>

8. Anexos

Quadro 8.1: Dados financeiros das lojas Meu Super

	Vendas Brutas Totais (u.m)	Vendas Brutas Fidelizadas (u.m)	Transações Totais (nº)	Transações Fidelizadas (nº)	Vendas Brutas Fidelizadas (%)	Transações Fidelizadas (%)
Março	5 852 201	2 980 969	956559	339027	50.9%	35.4%
Abril	5 778 204	3 092 240	934976	346090	53.5%	37.0%
Maio (13)	2 493 865	1 331 529	405048	147138	53.4%	36.3%

$$Densidade\ Populacional = \frac{Residentes}{Área\ (km^2)} \tag{8.1}$$

$$DDP = Residentes * \left(\frac{Área\ de\ cada\ mancha\ urbana}{Área\ Total} \right) \tag{8.2}$$

$$Deslocação\ a\ pé = \frac{(comprimento\ de\ cada\ troço * 60)}{3000} \tag{8.3}$$

Quadro 8.2: Matriz de Correlação

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
01	1	0.376	0.105	0.441	-0.031	-0.014	-0.055	-0.042	-0.078	-0.033	-0.011	-0.036	-0.041	-0.043	-0.032	-0.071	-0.053	-0.03	-0.077	-0.061	0.009	-0.043	-0.045	-0.07	-0.083	0.011	0.003	-0.042
02	0.376	1	0.368	0.855	-0.003	-0.023	0.003	-0.003	0.02	-0.009	0.009	-0.017	-0.004	-0.002	0.005	-0.012	0.014	-0.007	0.005	-0.041	0.048	-0.02	-0.008	-0.077	-0.024	0.019	0.026	0.048
03	0.105	0.368	1	0.356	-0.11	-0.077	-0.051	-0.1	-0.057	-0.108	-0.089	-0.104	-0.101	-0.099	-0.105	-0.093	-0.08	-0.109	-0.072	-0.098	-0.089	-0.109	-0.111	-0.098	-0.05	-0.036	-0.049	-0.055
04	0.441	0.855	0.356	1	-0.084	-0.094	-0.008	-0.085	-0.057	-0.089	-0.061	-0.096	-0.086	-0.084	-0.072	-0.084	-0.063	-0.088	-0.071	-0.109	-0.02	-0.099	-0.089	-0.139	-0.095	-0.051	-0.068	-0.038
05	-0.031	-0.003	-0.11	-0.084	1	0.767	-0.017	0.979	0.804	0.982	0.924	0.949	0.978	0.978	0.957	0.92	0.897	0.994	0.829	0.838	0.865	0.964	0.908	0.844	0.786	0.558	0.614	0.555
06	-0.014	-0.023	-0.077	-0.094	0.767	1	0.086	0.828	0.576	0.85	0.855	0.833	0.814	0.843	0.891	0.666	0.731	0.833	0.65	0.832	0.554	0.794	0.873	0.879	0.785	0.305	0.271	0.478
07	-0.055	0.003	-0.051	-0.008	-0.017	0.086	1	0.054	0.112	0.041	0.052	0.017	0.046	0.063	0.076	0.021	0.103	0.023	0.104	0.051	-0.016	-0.012	0.066	0.058	0.131	-0.148	-0.182	-0.32
08	-0.042	-0.003	-0.1	0.085	0.979	0.828	0.054	1	0.863	0.996	0.952	0.94	0.999	0.998	0.973	0.938	0.954	0.987	0.894	0.899	0.829	0.969	0.952	0.899	0.866	0.535	0.582	0.547
09	-0.078	0.02	-0.057	-0.057	0.804	0.576	0.112	0.863	1	0.818	0.729	0.669	0.877	0.843	0.756	0.938	0.961	0.795	0.986	0.742	0.749	0.833	0.793	0.704	0.803	0.492	0.578	0.46
10	-0.033	-0.009	-0.108	-0.089	0.982	0.85	0.041	0.996	0.818	1	0.958	0.95	0.992	0.997	0.983	0.918	0.929	0.993	0.859	0.903	0.818	0.97	0.957	0.907	0.855	0.519	0.559	0.543
11	-0.011	0.009	-0.089	-0.061	0.924	0.855	0.052	0.952	0.729	0.958	1	0.942	0.944	0.959	0.965	0.813	0.876	0.944	0.776	0.884	0.755	0.915	0.938	0.909	0.821	0.458	0.491	0.504
12	-0.036	-0.017	-0.104	-0.096	0.949	0.833	0.017	0.94	0.669	0.95	0.942	1	0.929	0.95	0.957	0.805	0.806	0.961	0.709	0.849	0.777	0.895	0.888	0.877	0.771	0.542	0.555	0.539
13	-0.041	-0.004	-0.101	-0.086	0.978	0.814	0.046	0.999	0.877	0.992	0.944	0.929	1	0.995	0.965	0.948	0.961	0.984	0.905	0.894	0.835	0.971	0.949	0.89	0.863	0.535	0.587	0.547
14	-0.043	-0.002	-0.099	-0.084	0.978	0.843	0.063	0.998	0.843	0.997	0.959	0.95	0.995	1	0.98	0.924	0.942	0.989	0.878	0.902	0.82	0.964	0.954	0.907	0.867	0.534	0.575	0.545
15	-0.032	0.005	-0.105	-0.072	0.957	0.891	0.076	0.973	0.756	0.983	0.965	0.957	0.965	0.98	1	0.847	0.887	0.98	0.808	0.878	0.778	0.944	0.954	0.907	0.843	0.463	0.487	0.505
16	-0.071	-0.012	-0.093	-0.084	0.92	0.666	0.021	0.938	0.938	0.918	0.813	0.805	0.948	0.924	0.847	1	0.951	0.909	0.937	0.836	0.806	0.93	0.872	0.798	0.818	0.559	0.629	0.53
17	-0.053	0.014	-0.08	-0.063	0.897	0.731	0.103	0.954	0.961	0.929	0.876	0.806	0.961	0.942	0.887	0.951	1	0.901	0.978	0.847	0.801	0.918	0.903	0.826	0.869	0.499	0.567	0.497
18	-0.03	-0.007	-0.109	-0.088	0.994	0.833	0.023	0.987	0.795	0.993	0.944	0.961	0.984	0.989	0.98	0.909	1	1	0.829	0.866	0.841	0.967	0.934	0.879	0.815	0.53	0.573	0.553
19	-0.077	0.005	-0.072	-0.071	0.829	0.65	0.104	0.894	0.986	0.859	0.776	0.709	0.905	0.878	0.808	0.937	0.978	0.829	1	0.789	0.735	0.86	0.844	0.765	0.858	0.482	0.57	0.477
20	-0.061	-0.041	-0.098	-0.109	0.838	0.832	0.051	0.899	0.742	0.903	0.884	0.849	0.894	0.902	0.878	0.836	0.847	0.866	0.789	1	0.551	0.875	0.944	0.963	0.912	0.453	0.471	0.481
21	0.009	0.048	-0.089	-0.02	0.865	0.554	-0.016	0.829	0.749	0.818	0.755	0.777	0.835	0.82	0.778	0.806	0.801	0.841	0.735	0.551	1	0.797	0.648	0.533	0.487	0.521	0.556	0.469
22	-0.043	-0.02	-0.109	-0.099	0.964	0.794	-0.012	0.969	0.833	0.97	0.915	0.895	0.971	0.964	0.944	0.93	0.918	0.967	0.86	0.875	0.797	1	0.948	0.862	0.801	0.484	0.527	0.513
23	-0.045	-0.008	-0.111	-0.089	0.908	0.873	0.066	0.952	0.792	0.957	0.938	0.888	0.949	0.954	0.954	0.872	0.903	0.934	0.844	0.944	0.648	0.948	1	0.952	0.906	0.416	0.46	0.487
24	-0.07	-0.077	-0.098	-0.139	0.844	0.879	0.058	0.899	0.704	0.907	0.909	0.877	0.89	0.907	0.907	0.798	0.826	0.879	0.765	0.963	0.533	0.862	0.952	1	0.925	0.402	0.465	0.511
25	-0.083	-0.024	-0.05	-0.095	0.786	0.785	0.131	0.866	0.803	0.855	0.821	0.771	0.863	0.867	0.843	0.818	0.869	0.815	0.858	0.912	0.487	0.801	0.906	0.925	1	0.435	0.506	0.491
26	0.011	0.019	-0.036	-0.051	0.558	0.305	-0.148	0.535	0.492	0.519	0.458	0.542	0.535	0.534	0.463	0.559	0.499	0.53	0.482	0.453	0.521	0.484	0.416	0.402	0.435	1	0.7	0.479
27	0.003	0.026	-0.049	-0.068	0.614	0.271	-0.182	0.582	0.578	0.559	0.491	0.555	0.587	0.575	0.487	0.629	0.567	0.573	0.57	0.471	0.556	0.527	0.46	0.465	0.506	0.7	1	0.706
28	-0.042	0.048	-0.055	-0.038	0.555	0.478	-0.32	0.547	0.46	0.543	0.504	0.539	0.547	0.545	0.505	0.53	0.497	0.553	0.477	0.481	0.469	0.513	0.487	0.511	0.491	0.479	0.706	1